HEALTH TECHNOLOGY ASSESSMENT

VOLUME 22 ISSUE 64 NOVEMBER 2018
ISSN 1366-5278

Prognostic models for identifying risk of poor outcome in people with acute ankle sprains: the SPRAINED development and external validation study

David J Keene, Michael M Schlüssel, Jacqueline Thompson, Daryl A Hagan, Mark A Williams, Christopher Byrne, Steve Goodacre, Matthew Cooke, Stephen Gwilym, Philip Hormbrey, Jennifer Bostock, Kirstie Haywood, David Wilson, Gary S Collins and Sarah E Lamb on behalf of the SPRAINED study group

NHS National Institute for Health Research

DOI 10.3310/hta22640

Prognostic models for identifying risk of poor outcome in people with acute ankle sprains: the SPRAINED development and external validation study

David J Keene,[1]* Michael M Schlüssel,[1] Jacqueline Thompson,[1] Daryl A Hagan,[1] Mark A Williams,[2] Christopher Byrne,[3] Steve Goodacre,[4] Matthew Cooke,[5] Stephen Gwilym,[1,6] Philip Hormbrey,[6] Jennifer Bostock,[7] Kirstie Haywood,[5] David Wilson,[6] Gary S Collins[1] and Sarah E Lamb[1,5] on behalf of the SPRAINED study group

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]Department of Sport, Health Sciences and Social Work, Oxford Brookes University, Oxford, UK
[3]Faculty of Health and Human Sciences, University of Plymouth, Plymouth, UK
[4]School of Health and Related Research, University of Sheffield, Sheffield, UK
[5]Warwick Medical School, University of Warwick, Coventry, UK
[6]Emergency Department, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[7]Patient and public involvement representative

*Corresponding author

Declared competing interests of authors: Steve Goodacre is a member of the Health Technology Assessment (HTA) Clinical Trials Board, HTA Elective and Emergency Specialist Care Methods Group, HTA Funding Boards Policy Group (formerly CSG), HTA IP Methods Group, HTA Post board funding teleconference and HTA Prioritisation Group. David Wilson declares personal fees from Oxford University. Gary S Collins is a member of the HTA Commissioning Board. Sarah E Lamb is Co-director of Oxford Clinical Trials Unit and Professor of Rehabilitation at Warwick Clinical Trials Unit; both receive funding from the National Institute of Health Research (NIHR). She is also a member of the HTA Additional Capacity Funding Board, HTA End of Life Care and Add on Studies, HTA Prioritisation Group and the HTA Trauma Board. Furthermore, she reports grants from the NIHR HTA programme during the conduct of this study.

Published November 2018
DOI: 10.3310/hta22640

This report should be referenced as follows:

Keene DJ, Schlüssel MM, Thompson J, Hagan DA, Williams MA, Byrne C, *et al*. Prognostic models for identifying risk of poor outcome in people with acute ankle sprains: the SPRAINED development and external validation study. *Health Technol Assess* 2018;**22**(64).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE, *Science Citation Index Expanded* (SciSearch®) and *Current Contents®*/ Clinical Medicine.

Health Technology Assessment

HTA/HTA TAR

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 4.513

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

The HTA programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

For more information about the HTA programme please visit the website: http://www.nets.nihr.ac.uk/programmes/hta

This report

The research reported in this issue of the journal was funded by the HTA programme as project number 13/19/06. The contractual start date was in November 2014. The draft report began editorial review in August 2017 and was accepted for publication in February 2018. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).

Abstract

Prognostic models for identifying risk of poor outcome in people with acute ankle sprains: the SPRAINED development and external validation study

David J Keene,[1]* Michael M Schlüssel,[1] Jacqueline Thompson,[1] Daryl A Hagan,[1] Mark A Williams,[2] Christopher Byrne,[3] Steve Goodacre,[4] Matthew Cooke,[5] Stephen Gwilym,[1,6] Philip Hormbrey,[6] Jennifer Bostock,[7] Kirstie Haywood,[5] David Wilson,[6] Gary S Collins[1] and Sarah E Lamb[1,5] on behalf of the SPRAINED study group

[1]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]Department of Sport, Health Sciences and Social Work, Oxford Brookes University, Oxford, UK
[3]Faculty of Health and Human Sciences, University of Plymouth, Plymouth, UK
[4]School of Health and Related Research, University of Sheffield, Sheffield, UK
[5]Warwick Medical School, University of Warwick, Coventry, UK
[6]Emergency Department, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[7]Patient and public involvement representative

*Corresponding author david.keene@ndorms.ox.ac.uk

Background: Ankle sprains are very common injuries. Although recovery can occur within weeks, around one-third of patients have longer-term problems.

Objectives: To develop and externally validate a prognostic model for identifying people at increased risk of poor outcome after an acute ankle sprain.

Design: Development of a prognostic model in a clinical trial cohort data set and external validation in a prospective cohort study.

Setting: Emergency departments (EDs) in the UK.

Participants: Adults with an acute ankle sprain (within 7 days of injury).

Sample size: There were 584 clinical trial participants in the development data set and 682 recruited for the external validation study.

Predictors: Candidate predictor variables were chosen based on availability in the clinical data set, clinical consensus, face validity, a systematic review of the literature, data quality and plausibility of predictiveness of the outcomes.

Main outcome measures: Models were developed to predict two composite outcomes representing poor outcome. Outcome 1 was the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence. Outcome 2 included the same symptoms as outcome 1, with the addition of recurrence of injury. Rates of poor outcome in the external data set

were lower than in the development data set, 7% versus 20% for outcome 1 and 16% versus 24% for outcome 2.

Analysis: Multiple imputation was used to handle missing data. Logistic regression models, together with multivariable fractional polynomials, were used to select variables and identify transformations of continuous predictors that best predicted the outcome based on a nominal alpha of 0.157, chosen to minimise overfitting. Predictive accuracy was evaluated by assessing model discrimination (*c*-statistic) and calibration (flexible calibration plot).

Results: (1) Performance of the prognostic models in development data set – the combined *c*-statistic for the outcome 1 model across the 50 imputed data sets was 0.74 [95% confidence interval (CI) 0.70 to 0.79], with good model calibration across the imputed data sets. The combined *c*-statistic for the outcome 2 model across the 50 imputed data sets was 0.70 (95% CI 0.65 to 0.74), with good model calibration across the imputed data sets. Updating these models, which used baseline data collected at the ED, with an additional variable at 4 weeks post injury (pain when bearing weight on the ankle) improved the discriminatory ability (*c*-statistic 0.77, 95% CI 0.73 to 0.82, for outcome 1 and 0.75, 95% CI 0.71 to 0.80, for outcome 2) and calibration of both models. (2) Performance of the models in the external data set – the combined *c*-statistic for the outcome 1 model across the 50 imputed data sets was 0.73 (95% CI 0.66 to 0.79), with a calibration plot intercept of –0.91 (95% CI –0.98 to 0.44) and slope of 1.13 (95% CI 0.76 to 1.50). The combined *c*-statistic for the outcome 2 model across the 50 imputed data sets was 0.63 (95% CI 0.58 to 0.69), with a calibration plot intercept of –0.25 (95% CI –0.27 to 0.11) and slope of 1.03 (95% CI 0.65 to 1.42). The updated models with the additional pain variable at 4 weeks had improved discriminatory ability over the baseline models but not better calibration.

Conclusions: The SPRAINED (Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department) prognostic models performed reasonably well, and showed benefit compared with not using any model; therefore, the models may assist clinical decision-making when managing and advising ankle sprain patients in the ED setting. The models use predictors that are simple to obtain.

Limitations: The data used were from a randomised controlled trial and so were not originally intended to fulfil the aim of developing prognostic models. However, the data set was the best available, including data on the symptoms and clinical events of interest.

Future work: Further model refinement, including recalibration or identifying additional predictors, may be required. The effect of implementing and using either model in clinical practice, in terms of acceptability and uptake by clinicians and on patient outcomes, should be investigated.

Trial registration: Current Controlled Trials ISRCTN12726986.

Funding: This project was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme and will be published in full in *Health Technology Assessment*; Vol. 22, No. 64. See the NIHR Journals Library website for further project information. Funding was also recieved from the NIHR Collaboration for Leadership in Applied Health Research, Care Oxford at Oxford Health NHS Foundation Trust, NIHR Biomedical Research Centre, Oxford, and the NIHR Fellowship programme.

Contents

List of tables

List of figures

List of boxes

List of abbreviations

AFS	Ankle Function Score
AIC	Akaike information criterion
BMI	body mass index
CAI	chronic ankle instability
CAST	Collaborative Ankle Support Trial
CDF	clinical data set form
CI	confidence interval
CRF	case report form
DCA	decision curve analysis
ED	emergency department
EPV	events per variable
EQ-5D	EuroQol-5 Dimensions
ESP	Extended Scope Practitioner
FAOS	Foot and Ankle Outcome Score
GCP	Good Clinical Practice
HeLEX	Centre for Health, Law and Emerging Technologies
HTA	Health Technology Assessment
ICF	informed consent form
LOWESS	locally weighted scatterplot smoothing
MAR	missing at random
MeSH	medical subject heading
MFP	multivariable fractional polynomial
MICE	multiple imputation by chained equations
mNGT	modified nominal group technique
MRI	magnetic resonance imaging
NIHR	National Institute for Health Research
OCTRU	Oxford Clinical Trials Research Unit
PPI	patient and public involvement
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUIPS	Quality In Prognosis Studies
RCT	randomised controlled trial
REC	Research Ethics Committee
SD	standard deviation
SMG	Study Management Group
SPRAINED	Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department
SSC	Study Steering Committee

Plain English summary

Sprains of the ankle joint ligaments are very common injuries. Most people recover within a few weeks but up to one in three people have a poor outcome. A poor outcome includes problems such as ongoing pain, difficulties moving about, lack of confidence and further sprains. It is challenging to work out who will recover and who will not because, when people come into emergency departments (EDs) for assessment, the ankle is often so sore that the patient cannot tolerate a thorough examination.

We developed a tool to help predict who is at greater risk of a poor outcome. A tool like this would be useful as it would have the potential to assist clinical decision-making and could help identify the people with an acute ankle sprain who could benefit from rehabilitation and monitoring.

The tool takes into account participant characteristics, such as age, and injury characteristics, such as the severity of pain reported. The tool had good accuracy among a group of participants who had been involved in a clinical trial. To see how the tool performed in another group of participants, we recruited 682 participants from 10 EDs in the UK. We collected information on the participant and injury characteristics when the participant attended the ED and again 9 months afterwards. The research indicated that the tool has moderate ability to predict what will happen in the future. There are limitations to the accuracy of the predictions of the tool. However, our analyses suggest that using the tool is better than the scenario of not using a tool to identify people at risk of a poor outcome after ankle sprain.

To make use of the tool in clinical settings, it would benefit from being set up on a web-based application or a similar mobile platform to enable clinicians to enter information about a patient and obtain a calculated risk score. The prediction tool could also be improved by further research to see how well it performs in routine clinical care and in other settings.

Scientific summary

Background

Ankle sprains are one of the most common musculoskeletal injuries. Although recovery can occur within weeks, up to one-third of patients still have problems with their ankle at 1 year post injury. In the acute phase there is no reliable way of establishing which patients are at risk of having a poor outcome.

Objectives

To develop prognostic models to be used in an acute setting to identify people at increased risk of poor outcome following an acute ankle sprain, and to evaluate the performance of these prognostic models in a prospective external validation study.

Methods

Research programme

A systematic review of prognostic factors for poor outcome after ankle sprain was conducted, followed by an expert consensus meeting, then development of prognostic models and external validation using data from a prospective observational cohort study.

Systematic review

The review was registered on the PROSPERO database: CRD42014014471. Electronic databases were searched [Allied and Complementary Database (AMED), EMBASE, PsycINFO, Cumulative Index to Nursing and Allied Health Literature (CINAHL) and SPORTDiscus, PubMed, Cochrane Register of Clinical Trials, Physiotherapy Evidence Database (PEDro)]. Studies that had participants with acute ankle sprain, a longitudinal design and assessment of at least one baseline prognostic factor were included. Eligibility assessments, data extraction and risk-of-bias assessments [using the Quality In Prognosis Studies (QUIPS) tool] were completed by two independent reviewers. A narrative synthesis was conducted.

Consensus meeting

A range of key stakeholders involved in ankle sprain care and research in the UK NHS, including patient and public representatives, health-care professionals and clinical researchers, were invited to a 1-day consensus meeting.

A modified nominal group technique (mNGT) was used to facilitate the consensus process. The participants were divided into three groups (participants were pre-assigned to groups to ensure that there was a mixture of clinicians, researchers and patient representatives in each group) and were asked to rank important prognostic factors, some of which were nominated in the pre-meeting questionnaire. Discussions were immediately followed by a plenary session to report results of the group discussions to the entire group. A final session comprised a voting process, in which each participant indicated whether or not each factor should be included in the prognostic model. The number of votes allowed was limited to 10 per individual. This was completed independently on paper questionnaires. Factors with $\geq 70\%$ agreement across participants were considered critically important and eligible for inclusion in the validation study.

Development of the models

Data sources

Individual participant data from the existing Collaborative Ankle Support Trial (CAST) database were used to develop two prognostic models for poor outcome after ankle sprain. CAST was a pragmatic, multicentre, randomised controlled trial (RCT), with blinded assessment of the outcome, designed to estimate the clinical effectiveness and cost-effectiveness of three different types of mechanical ankle support [Aircast® ankle brace (DJO Incorporated, Vista, CA, USA), Bledsoe® boot (Bledsoe Boot Systems, Grand Prairie, TX, USA) or 10-day below-knee cast] for the initial management of severe ankle sprain (defined as an injury of grade 2 or 3, without fracture) compared with a double-layer tubular compression bandage.

The trial population comprised 584 individuals aged ≥ 16 years attending emergency departments (EDs) in the UK with an ankle sprain and an inability to fully bear weight on the injured ankle at the time of presentation to the ED and at their review clinic appointment (the trial's baseline assessment). People were excluded if they presented with an ankle fracture (apart from a flake fractures of < 2 mm), any other recent fracture, any contraindication to any of the four arms of the trial, poor skin viability preventing splinting or casting, or if their injury occurred > 7 days before the first presentation at the recruiting ED. Participants were followed up at 1, 3 and 9 months after randomisation.

Candidate predictors

Twenty-three candidate predictor variables collected during the enrolment and baseline assessments of CAST were examined; all of these variables came under the following domains: age, sex, pain, previous injury, ankle stability tests, weight-bearing ability and severity of presenting clinical signs and symptoms. These candidate predictor variables were chosen based on clinical consensus, face validity, systematic review of the literature, data quality and whether or not they were plausibly predictive of the outcomes.

Outcomes

The first prognostic model was developed to predict a composite outcome representing the presence of at least one of the following symptoms at 9 months post injury: persistent pain, functional difficulty or lack of confidence (outcome 1).

The second model was developed to predict a composite outcome representing the presence of at least one of the following symptoms or clinical events at 9 months post injury: persistent pain, functional difficulty, lack of confidence or recurrence of injury (outcome 2).

Sample size

Based on the CAST data set, between 20% (116/584) and 24% (140/584) of people attending an ED for an acute ankle sprain experienced a poor outcome at 9 months. As this was the first study aiming to produce prognostic models to predict poor outcome after ankle sprain, we relaxed the recommendation of five events per variable (EPV) for the number of variables in a logistic regression model. We included 23 candidate predictors (with a total of 35 degrees of freedom) in the full model, which meant an EPV ratio of approximately 3 (116/35) and 4 (140/35) for outcomes 1 and 2, respectively.

Analysis

Multiple imputation was used to handle missing data, with 50 imputed data sets created. Based on a logistic regression model, multivariable fractional polynomials (MFPs) were used to select variables and identify transformations of continuous variables that best predicted the outcome. Inclusion of predictors in the final models was based on a nominal alpha of 0.157 (equivalent to the Akaike information criterion) to reduce the risk of overfitting. Shrinkage of the regression coefficients and intercepts was performed based on heuristic shrinkage factors to correct for optimism. Predictive accuracy of the models was evaluated by assessing model discrimination (quantified by the *c*-statistic) and model calibration (flexible calibration plot).

External validation of the model

A prospective cohort study recruited people with acute ankle sprain attending one of 10 NHS EDs across England over a period of 9 months (July 2015–March 2016). There was no planned treatment allocation, as in a RCT, and EDs provided usual care in accordance with local protocols. Data collection took place at the time of a participant's presentation to any of the study recruiting sites (baseline) and subsequently at 4 weeks and 4 and 9 months after the initial injury. People aged ≥ 16 years with an acute ankle sprain (of < 7 days' duration) of any severity were invited to take part in the study. People with an ankle fracture (except a flake fracture of < 2 mm) or other recent (< 3 months) lower limb fracture were excluded. During this part of the study, a pilot of dynamic consent was also included in the later stages of recruitment. This gave participants an opportunity to use a website to interact with study information and update their preferences.

Results

Systematic review

Searches identified 4173 reports, with eight reports identified from additional sources. Thirty-six reports were assessed in full-text screening and nine studies were included in the review.

One study was judged to be at low risk of bias, five at moderate risk of bias and three studies at high risk of bias. Incomplete and/or inadequate reporting standards were a common issue; for example, it was difficult to determine if prognostic factors were eliminated because of statistical reasons or poor clinical utility. None of the studies reported on performance of the prognostic models using methods to assess internal or external validation. Across the included studies, a wide range of prognostic factors was investigated. The prognostic factors that were analysed varied considerably between studies, with no common framing across the studies. The identified studies and risk-of-bias assessments were summarised to those attending the consensus meeting.

Consensus meeting

The consensus meeting was attended by 30 participants. The final consensus voting identified eight baseline factors that were deemed critical for the identification of people likely to have a poor recovery. These factors spanned pre-injury, sociodemographic, psychosocial and clinical assessment factors, encompassing a holistic biopsychosocial model of recovery. These factors were included in the data collection at baseline for the prospective observational study.

Performance of the prognostic models in development data set

The first model predicted the presence of persistent pain, functional difficulty or lack of confidence at 9 months and comprised age, body mass index, pain when resting, pain when bearing weight, number of days from injury to assessment, whether or not the injury is a recurrent sprain and the ability to bear any weight on the injured ankle (outcome 1). The apparent performance on a complete-case analysis of the CAST data set showed a *c*-statistic of 0.82 [95% confidence interval (CI) 0.75 to 0.89]. The combined *c*-statistic across the 50 imputed data sets was 0.74 (95% CI 0.70 to 0.79), with good model calibration.

The second model predicted the presence of either persistent pain, functional difficulty, lack of confidence or recurrence of injury at 9 months and comprised pain when resting, pain when bearing weight, days from injury to assessment, ability to bear any weight on the injured ankle and whether or not the injury is a recurrent sprain (outcome 2). The apparent performance on a complete-case analysis of the CAST data set showed a *c*-statistic of 0.73 (95% CI 0.66 to 0.81). The combined *c*-statistic across the 50 imputed data sets was 0.70 (95% CI 0.65 to 0.74), with good model calibration.

Updating these models, which used baseline data collected at the ED, with an additional variable at 4 weeks after the injury (pain when bearing weight on the ankle), improved the predictions of the models when compared, using decision curve analysis plots.

A substudy to pilot dynamic consent recruited 22 participants in the later phase of the prospective cohort study. Eight participants accessed their dynamic consent online web page and none changed his/her consent decisions during the study.

Performance of the models in the external data set

Discrimination of the model for outcome 1 was similar to that observed in the development data set (combined *c*-statistic across the 50 imputed data sets = 0.73, 95% CI 0.66 to 0.79), but calibration was poor (combined calibration plot intercept = –0.91, 95% CI –1.18 to –0.65, and slope = 1.13, 95% CI 0.76 to 1.50). For the outcome 2 model, the combined *c*-statistic across the 50 imputed data sets was 0.63 (95% CI 0.58 to 0.69), the calibration plot intercept was –0.25 (95% CI –0.44 to –0.06) and the slope was 1.03 (95% CI 0.65 to 1.42). Discrimination of the updated model for outcome 1 was better (combined *c*-statistic = 0.78, 95% CI 0.72 to 0.84), but calibration did not improve substantially (combined calibration plot intercept = –0.62, 95% CI –0.89 to –0.34, and slope = 1.17, 95% CI 0.86 to 1.48). The combined *c*-statistic for the updated model for outcome 2 was 0.64 (95% CI 0.59 to 0.69), the calibration plot intercept was 0.12 (95% CI –0.07 to –0.32) and slope was 0.68 (95% CI 0.46 to 0.91). Finally, model performance was not better for the subgroup of participants with more severe injuries (ankle sprains of grade 2 or 3). All models were recalibrated (i.e. had their regression coefficients and intercepts re-estimated) using the external validation data set.

A substudy to pilot dynamic consent recruited 22 participants in the later phase of the prospective cohort study. Eight participants accessed their dynamic consent online web page and none changed his/her consent decisions during the study.

Conclusions

Both models and their updates provided good predictions of poor outcome for people with acute ankle sprain on the population used in their derivation. There was a slight decrease in model discrimination for both models when evaluated in a prospectively collected external validation cohort. The models predicting presence of persistent pain, functional difficulty, lack of confidence or recurrence of injury showed good calibration, whereas there was miscalibration of the model predicting persistent pain, functional difficulty or lack of confidence. Recalibration of the models may be required to improve the accuracy of the predicted risks in other populations (within and outside the UK).

Implications for health care

The SPRAINED (Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department) study prognostic models performed reasonably well and showed benefit when compared with not using any model (i.e. consider all patients to be at a high risk of poor outcome); therefore, the models may assist clinical decision-making when assessing and advising people with ankle sprains in the ED setting and when deciding on ongoing management. The models benefit from using predictors that are simple to obtain during routine clinical assessment.

Recommendations for research

Further research to evaluate the performance of the models in other settings is recommended. Further refinement of the models, including external validation of the recalibrated models or identifying additional predictors, may be required. The impact of implementing and using either model in clinical practice, in terms of acceptability and uptake by ED staff, and their impact on patient outcomes, should be investigated.

Trial registration

This trial is registered as ISRCTN12726986.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research. Funding was also recieved from the NIHR Collaboration for Leadership in Applied Health Research, Care Oxford at Oxford Health NHS Foundation Trust, NIHR Biomedical Research Centre, Oxford and the NIHR Fellowship programme.

Chapter 1 Introduction

Background

Incidence and costs

Ankle sprains are one of the most common musculoskeletal injuries. Between 3% and 5% of people who attend an emergency department (ED) in the UK do so as a result of sustaining a sprained ankle.[1] The vast majority of sprains are of the lateral (outside) ligaments, and vary from minor stretching (grade 1) to a complete tear (grade 3).[2] Recent systematic reviews[3,4] conclude that ≈30% of people still have problems with their injury 1 year after an ankle sprain, depending on the outcome measured and, perhaps more importantly, the sampling frame. Many studies are restrictive in their sampling frame, either concentrating on elite athletes or excluding younger and older people. Studies also have variable inception and follow-up points, which further complicates interpretation. A large multicentre randomised controlled trial (RCT) conducted in EDs in the UK reported an estimated prevalence of poor outcome of 30% at 9 months.[5] Other studies agree that recovery plateaus at around 9 months, and that residual disability after this point is likely to be persistent.[6] One potential consequence of ankle sprain, chronic ankle instability (CAI), is implicated in the development of ankle osteoarthritis, even without an acute osteochondral lesion.[7]

Usual clinical pathway

Assessment of the injury in the acute phase is challenging as the ankle is often so swollen and painful that it cannot easily be examined. Most people are advised to rest, to elevate the ankle and to apply ice and compression; crutches are often issued if bearing weight is difficult. The Ottawa guidance[8] can be used to reduce the requirement for imaging without missing significant fractures. If clinicians are concerned about the degree of injury, most health-care providers operate a system of review within 1 week in a trauma clinic or equivalent injury service. This time frame allows some resolution of swelling, and greater certainty in ascertainment of injury severity and presence of other significant mechanical derangement.[9] Treatment options at this stage include further watchful waiting, diagnostics, intensive physiotherapy and immobilisation. Surgery may be considered at this stage, although most centres would initiate a test of conservative management first. We have previously published a survey of practice,[1] which remains a reasonable reflection of current management in the UK.

Value of a prognostic model

In this report we utilise the terms recommended in the Prognosis Research Strategy (PROGRESS)[10–12] framework to describe the different types of prognostic research. A prognostic factor is '. . . any measure that, among people with a given health condition (that is, a start point), is associated with a subsequent clinical outcome (an endpoint)'.[12] A prognostic model is '. . . a formal combination of multiple predictors from which risks of a specific endpoint can be calculated for individual patients'.[10]

A prognostic model is advised to identify people likely to experience poor outcome after ankle sprain. There are several ways in which better prognostic information could yield benefit to the NHS and to patients. The first way would be the ability to decide whether or not an early review is merited and avoid unnecessary appointments. The second way would be the ability to target treatments and diagnostics more effectively and earlier in the recovery pathway. Finally, it could offer reassurance that people with ankle sprains who are not followed up are likely to be on a positive recovery trajectory. The large number of people who sustain an ankle sprain is a key issue for management; cost savings will accrue if treatments are more efficiently targeted. Any prognostic model needs to be simple to complete in the ED, ideally administered in a single assessment.

Requirements of a prognostic model

To be considered useful, a prognostic model should be clinically meaningful, accurate (well-calibrated with good discrimination) and generalisable (have been evaluated on a separate data set, referred to as external validation). Many prognostic models are developed using data sets that are too small, are not sufficiently generalisable, have questionable methodological quality (in particular, no or limited evaluation of predictive accuracy) and use inadequate statistical methods.[10–12] Other issues in developing a prognostic model are variable selection, handling of missing data, timing and method (self-report vs. clinical examination).

Existing prognostic models

Hiller *et al.*[13] authored a systematic review of factors associated with the risk of sustaining an ankle sprain, but there are few studies evaluating the risk of poor recovery after the injury. Other than recurrent sprain, few studies of post-injury recovery have considered wider predispositional factors. In 2008, van Rijn *et al.*[3] published a systematic review of the clinical course and prognostic factors for recovery following ankle sprain. They found just one eligible study,[14] which concluded that a high level of sports activity was a prognostic factor for residual symptoms ($n = 150$).

To the best of our knowledge, there are no externally validated prognostic models for acute ankle sprain (see *Chapter 3*). Prognostic model studies to date are of limited generalisability because of highly selective patient populations (e.g. exclusion of some of the more severe types of injury, exclusion of older people and/or sole inclusion of athletic/military populations). We identified only one study that was judged as being of high methodological quality, but a limited number of candidate prognostic factors were assessed.[15] Therefore, development of a new prognostic model – by using robust methods, considering a range of plausible prognostic factors and conducting an external validation – is advisable.

Polzer *et al.*[4] developed a prognostic algorithm and treatment pathway, but substantial sections were based on expert judgements. A robustly developed and validated prognostic model could help better target treatment and improve outcomes for people who have an ankle sprain.[10] There are treatment options available for people who have poor prognosis. The treatment with the most solid evidence base is physiotherapy.[16] Other options include surgical reconstruction of ligaments.[17]

Aim of the SPRAINED study

The aim of the Synthesising a clinical Prognostic Rule for Ankle Injuries in the Emergency Department (SPRAINED) study was to develop and validate a prognostic model for use in EDs for people with acute ankle sprain in order to identify those for whom recovery may be substantially prolonged or incomplete.

Chapter 2 Overview of methods

The development of a prognostic model for ankle sprains required a research programme that was conducted in two stages and used a variety of research methods. In order to facilitate an understanding of the development and validation of the prognostic model, the methods used across the research programme are outlined in this chapter. Full descriptions of the methods for the different stages of the research are contained in the following chapters.

Summary of study design

The SPRAINED study had two stages, summarised in *Figure 1*.

Systematic review of the literature

A systematic review was conducted to identify prognostic factors of poor outcome following acute ankle sprain to identify variables that could be considered from the array available in the data set described below (see *Developing a multivariable prognostic model from the CAST data set*) and in the external validation study (see *External validation of the prognostic model in a prospective observational cohort study*).

Expert consensus process

A modified nominal group technique (mNGT) was used to gain consensus and information on preferences. Briefing papers containing lay summaries of the preliminary modelling elements completed and prognostic factors identified in the systematic review were prepared and circulated to clinicians, patient and public representatives and clinical researchers. The consensus element was achieved through a face-to-face meeting, at which small groups were facilitated to answer a prespecified set of questions. Two steps were used in this process, the first one for identification of issues and general discussion, and the second for resolution and consensus.

Developing a multivariable prognostic model from the CAST data set

The Collaborative Ankle Support Trial (CAST) is, to date, the largest registered RCT of interventions for moderate to severe ankle sprains worldwide ($n = 584$ participants).[18] Data were collected on a large number of candidate prognostic factors, including those identified as potentially important by clinical

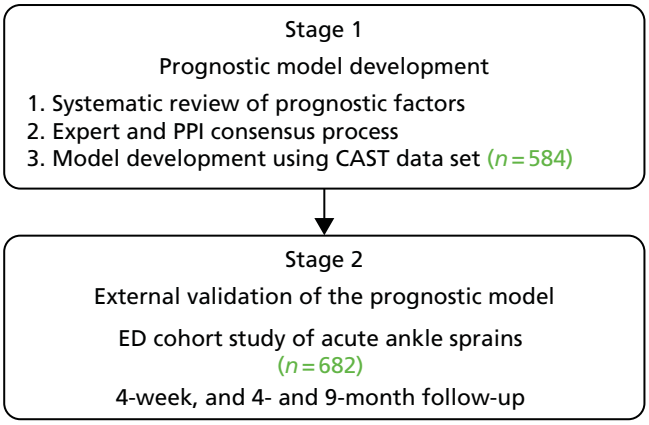


FIGURE 1 Stages of the SPRAINED study. PPI, patient and public involvement.

guidelines and consensus, and in previous multivariable analyses. The central research team had access to data at ED presentation, 2 to 3 days later, then at 1, 3 and 9 months after randomisation. Candidate prognostic factors were identified and included in multivariable models.

External validation of the prognostic model in a prospective observational cohort study

We conducted a prospective observation study of 682 participants across 10 EDs in England between 20 July 2015 and 17 March 2016. In this final part of the research, the prognostic model developed in the earlier work was externally validated and recalibrated. A baseline pro forma was used to obtain participant and clinical data on the candidate predictor variables, completed by the ED clinician at initial attendance. Follow-up data were collected from participants at 4 and 9 months via telephone, postal or online questionnaires, and captured persistent symptoms, the validated Foot and Ankle Outcome Score (FAOS),[19] health service resource use and health-related quality of life, measured using the EuroQol-5 Dimensions, three-level version (EQ-5D-3L).[20] An overview of this part of the study is contained in *Figure 2*. Data collected at baseline and 4 weeks after the injury were minimal, including mainly information on the predictors selected to compose the prognostic models developed for the two outcomes of interest. Data were also collected on a few baseline candidate predictors not present in the CAST data set to determine whether or not the prognostic validity of the models could be improved by the addition of this extra information.

Pilot of substudy of dynamic consent

Towards the end of recruitment for the external validation study, participants were offered the opportunity to join a dynamic consent pilot study. This gave participants an opportunity to use a website to interact with study information and update their preferences. Details of this substudy can be found in *Appendix 1*.

Patient and public involvement

The SPRAINED study recruited four patient and public involvement (PPI) representatives from a process of open advertisement on the People in Research website,[21] South Central Research Design Service e-bulletin, and the John Radcliffe Hospital ED in Oxford. Our appointed PPI representatives had experienced an ankle sprain and accessed NHS ED services. One representative agreed to be the PPI lead representative and is a co-applicant.

In order to develop and refine our application, we held a programme development meeting with our PPI representatives. Our representatives reviewed and contributed to ideas and provided feedback on our programmes of work, including who the team should consist of, the experience of service use from the PPI perspective, the relevance of our proposed outcomes, the acceptability of the research methods and the role of PPI input in developing and guiding the full application and research programme. We sought input on what were important outcomes and these influenced the make-up of our composite outcome measure.

The PPI representatives were involved in piloting the pre-consensus meeting questionnaire and participated in the consensus meeting. We also had input from the lead PPI representative on interpretation of the results and in planning dissemination during a Study Management Group (SMG) meeting; they were involved in reviewing the report.

Ethics approval and monitoring

Ethics approval for the SPRAINED study was given by the National Research Ethics Committee (REC) (London – Chelsea), REC number 15/LO/0538, on 10 April 2015. This trial was conducted in accordance with the ethics principles that have their origin in the Declaration of Helsinki[22] and that are consistent with Good Clinical

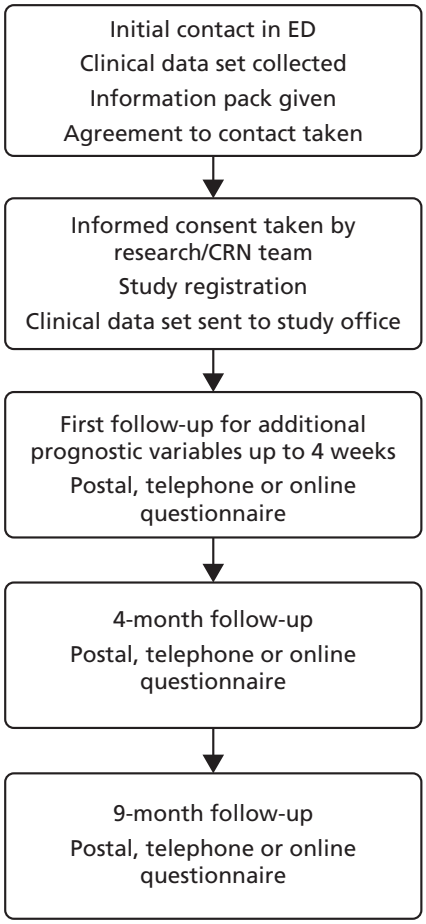


FIGURE 2 Flow chart of the SPRAINED cohort study. CRN, clinical research network.

Practice (GCP)[23] and the applicable requirements as stated in the UK Framework for Health and Social Care Research.[24] The sponsor of the study (University of Oxford) reviewed study documents before ethics submission.

The Oxford Clinical Trials Research Unit (OCTRU) assisted collaborating sites in obtaining the necessary approvals to allow the study to take place within their NHS trusts. The study was monitored and audited in accordance with the current approved protocol, GCP, relevant regulations and standard operating procedures. A monitoring plan was developed in accordance with OCTRU's standard operating procedures.

Study Steering Committee

The Study Steering Committee (SSC) provided overall supervision of the study on the behalf of the funder and was chaired by an independent member. The SSC abided by the OCTRU Standard Operating Procedure (accredited by the UK Clinical Research Collaboration Clinical Trials Unit registration process) and SSC charter. The SSC monitored study progress and advised on scientific credibility.

Study Management Group

The SMG was made up of SPRAINED study investigators and staff working on the project within OCTRU and the Critical Care, Trauma and Rehabilitation Trials Group. This group oversaw the day-to-day running of the trial and met regularly.

Reporting

The chief investigator submitted progress reports throughout the study period to the REC, host organisation and sponsor.

The description of the development and external validation of the two models followed the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis (TRIPOD) statement.[25]

A peer-reviewed journal manuscript was published to facilitate dissemination of the SPRAINED study prognostic model.[26]

Summary of changes to the study protocol and analysis plan

The changes to the study protocol are summarised in *Table 1*. The planned analysis was refined during the programme of research in line with methodological developments and in response to the findings between the development and external validation stages of the study. These refinements included the following:

- The primary outcome to represent 'poor outcome' after ankle sprain was clarified and prespecified in the analysis plan. This was as a result of the development of the research, considering the current literature and expert and PPI input. The final definitions were two different combinations of clinical features reported 9 months after injury:
  - Outcome 1 was the presence of at least one of the following symptoms at 9 months after injury – persistent pain, functional difficulty or lack of confidence.
  - Outcome 2 included the same symptoms as outcome 1 with the addition of recurrence of injury.
- Net reclassification improvement and integrated discrimination improvement were not carried out; instead a decision curve analysis (DCA) was undertaken (see *Chapter 5*).
- Decision curve analysis was not used to investigate the incremental value of a multivariable model with additional predictors not present in the development phase, as these predictors never reached that stage (see *Chapter 6*, *Model recalibration*).
- More than 15 candidate predictors were chosen for inclusion in the multivariable logistic regression models (see *Chapter 5*, *Sample size considerations* and *Data modelling*).
- The predictors selected for the final multivariable model were those meeting the threshold of $p < 0.157$ [equivalent to Akaike information criterion (AIC)] instead of backwards elimination with $p < 0.2$ as stopping rule, to minimise overfitting (see *Chapter 5*, *Data modelling*).
- Internal validation using bootstrapping was not done (not being possible without suppressing one or more of the strategies used to prevent overfitting). Instead, the heuristic shrinkage factors for each developed model were estimated and were used to correct intercepts and beta coefficients for optimism (see *Chapter 5*, *Assessment of model performance* and *Shrinkage*).
- Model presentation was not simplified to a scoring system. The final models developed were fairly simple, with only a few predictors commonly screened in clinical routine, so, instead, the equations with corresponding regression coefficients and intercepts were presented.

TABLE 1 Changes to the protocol during the study by version number

Amendment number	Protocol version number	Date issued	Details of changes made
1	2.0	11 November 2015	Added information on dynamic consent bolt-on study
2	3.0	3 March 2016	Clarification that follow-up time points are from study registration
3	4.0	28 July 2016	Addition of electronic/online methods of data collection taking place for all follow-up time points

Chapter 3 Systematic review

Introduction

A systematic review of prognostic factors for poor outcome following acute ankle sprain was conducted with the aim of identifying candidate variables that could be considered in the SPRAINED study. In this chapter, the methods, results and key findings of the systematic review that contributed to the development of the prognostic model are detailed.

Methods

The review protocol was registered on PROSPERO.[27]

Search strategy

Searches of the following electronic databases were conducted from inception to September 2016: Allied and Complementary Database (AMED), EMBASE, PsycINFO (via Ovid), Cumulative Index to Nursing and Allied Health Literature (CINAHL) and SPORTDiscus (via EBSCO*host*), PubMed and Cochrane Register of Clinical Trials. Relevant medical subject heading (MeSH) terms were used when appropriate in these databases. Search strings containing terms for the health condition or body region were used in Physiotherapy Evidence Database (PEDro), International Foot and Ankle Biomechanics, International Ankle Symposium and OpenGrey. No language restrictions were applied and the reference lists of included studies were screened for potentially relevant studies. The search strategy is available in *Appendix 2*.

Eligibility criteria

Studies were eligible for inclusion if they had all of the following factors:

- a sample, or a separately analysed subgroup, with a clinical diagnosis of acute (≤ 7 days from injury to assessment) lateral ankle ligament sprain
- a longitudinal design, with at least one follow-up time point
- statistical assessment of at least one baseline prognostic factor on recovery outcomes.

Excluded studies were those that included participants with ankle fracture (excluding flake fracture of < 2 mm) and other recent (< 3 months since injury) lower limb injuries.

Data extraction

Titles and abstract were screened by two members of the review team (JT, CB or MAW). The Ouzzani *et al.*[28] systematic review web application was used to manage screening. Full-text articles for potentially eligible records were independently reviewed by two of three reviewers (JT, CB or MAW). Data extraction and risk-of-bias assessments were completed independently by two reviewers (JT and CB). Discrepancies between reviewers decisions were resolved by discussion, or in consultation with a third reviewer (MMS or DJK).

Risk-of-bias assessment

Study quality was assessed using the Quality In Prognosis Studies (QUIPS) tool,[29] which considers the six following domains of validity and risk of bias in prognostic factor studies:

1. study participation
2. study attrition
3. prognostic factor measurement
4. confounding measurement and account

5. outcome measurement
6. analysis and reporting.

Data synthesis and reporting

A narrative synthesis was conducted, meta-analysis being considered inappropriate because of heterogeneity in the prognostic factors, outcome measures and follow-up durations and limited number of studies. Follow-up time points from injury were grouped as short term (≤ 8 weeks), medium term (≤ 4 months) and long term (> 4 months).

Results

Searches identified 4173 reports, with eight reports identified from additional sources. *Figure 3* shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram. There were 36 reports assessed in full-text screening. Of these, 27 were excluded; the remaining nine studies were included in the review.[15,30–37]

Study characteristics

Table 2 illustrates the characteristics of the nine included studies. Six studies[30–33,35,37] were prospective cohorts and three[15,34,36] were retrospective analyses of RCTs. Three studies were based in the Netherlands,[30,34,35] three in the USA,[31,32,37] and one each in England,[15] Northern Ireland[36] and Germany.[33] The median participant sample size was 33 (range 20–553 participants), and follow-up data ranged from 1 day to 12 months after injury. Three studies[31,32,37] recruited high school or university athletes; the remainder were based in primary or secondary care.

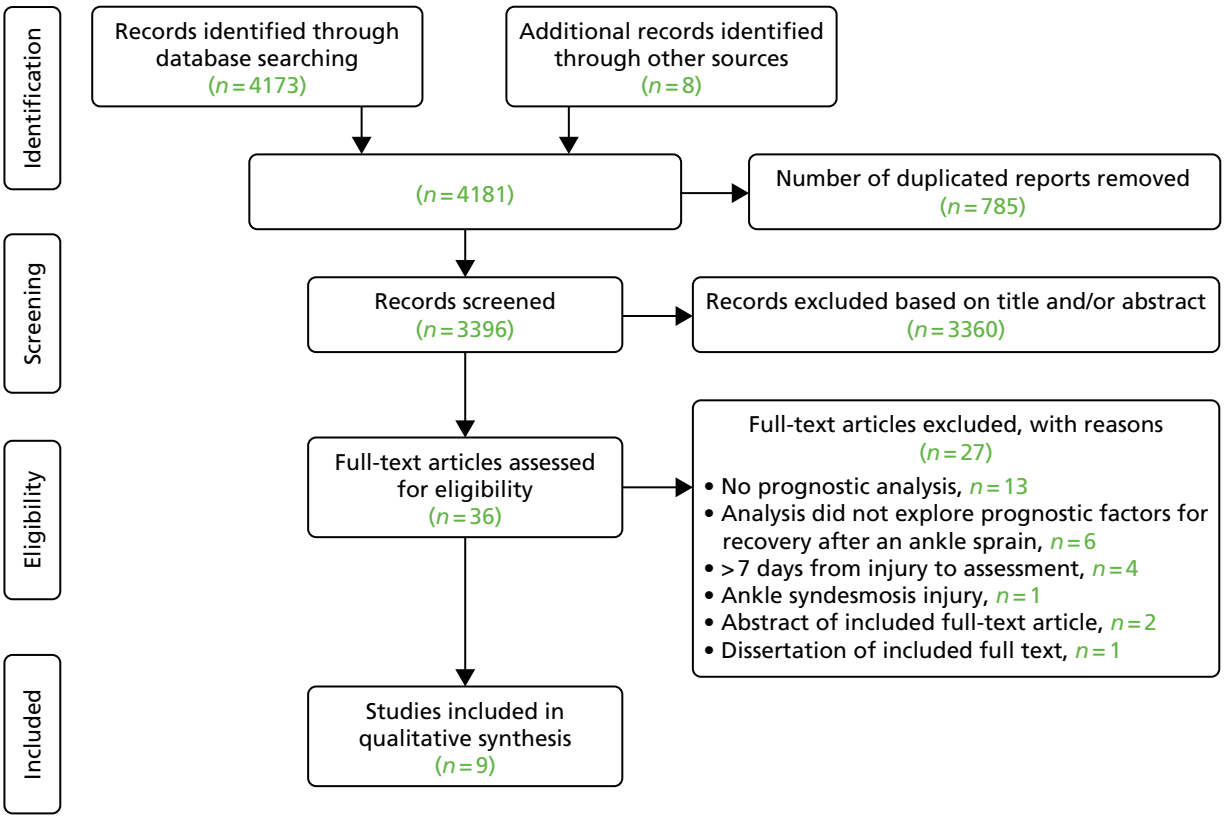


FIGURE 3 The PRISMA flow diagram.

TABLE 2 Key characteristics of included studies

Study	Study characteristics						
	Design	Setting	Sample size (*n*)	Sample characteristics	Time from injury to assessment	Injury severity	Follow-up
de Bie *et al.*[30]	Prospective cohort	The Netherlands: • 1 × hospital FAD	• 35 at baseline • 33 at 2 weeks • 31 at 4 weeks	General population 22 male; 13 female Average age 28 years, SD 10 years; range 13–59 years	NR	NR	• 2 weeks • 4 weeks
Wilson and Gansneder[31]	Prospective cohort	USA: • 1 × university	• 24 at baseline • 21 at follow-up	Athletes 13 male; 8 female Average age 20 years, SD 2 years	67.8 hours (SD 15.2 hours)	Grades I and II	11.9 days (SD 6.6 days)
Cross *et al.*[32]	Prospective cohort	USA: • 1 × university	• 20 at baseline • 20 at follow-up	Athletes 7 male; 13 female Average age 19 years, SD 1 year; range 18–21 years	≤ 24 hours	NR	14.7 days (SD 8.8 days), range 3–40 days
Akacha *et al.*[15]	Retrospective analysis	England: • 8 × hospital ED	• 584 at baseline • 553 at 4 weeks, 12 weeks and 9 months	General population 321 male; 232 female Average age 30 years, SD 11 years; range 16–72 years	≤ 7 days	Severe (NWB status at 3 days)	• 4 weeks • 12 weeks • 9 months
Langner *et al.*[33]	Prospective cohort	Germany: • 1 × hospital ED	• 38 at baseline • 26 at 6 months • NR at 12 months	General population 18 male; 20 female Average age 38 years, SD 13 years; range 20–75 years	< 24 hours	ATFL grade I (27%), grade II (27%) and grade III (46%)	• 6 months • 12 months

continued

TABLE 2 Key characteristics of included studies (*continued*)

	Study characteristics						
Study	**Design**	**Setting**	**Sample size (*n*)**	**Sample characteristics**	**Time from injury to assessment**	**Injury severity**	**Follow-up**
van Middelkoop *et al.*[34]	Retrospective analysis	The Netherlands: • 32 × general practice primary care • 1 × hospital ED	• 102 at baseline • 95 at 3 months • 80 at 12 months	General population 59 male; 43 female Average age 37 years, SD 12 years; range 18–60 years	≤ 7 days	Mild (42%), moderate or severe (44%), unknown (14%)	• 3 months • 12 months
van der Wees *et al.*[35]	Prospective cohort	The Netherlands: • 20 × primary care physiotherapists	• 107 at baseline • 33 at 2 weeks	General population 65 male; 42 female Average age 32 years, SD 14 years	• 8.7 days (SD 8.9 days) • ≤ 5 days for *n* = 53 participants • > 5 days for *n* = 54 participants	Light (50%), severe (50%)	2 weeks
O'Connor *et al.*[36]	Retrospective analysis	Northern Ireland: • 1 × hospital ED • 1 × university sports injury clinic	• 101 at baseline • NR at 4 weeks • 85 at 4 months	General population, athletes 69 male; 31 female Average age 27 years, SD 10 years; range 16–58 years	• < 7 days • 40 hours (SD 36 hours)	Grade I (26%), grade II (63%), grade II+ (11%)	• 4 weeks • 4 months
Medina McKeon *et al.*[37]	Prospective cohort	USA: • 7 × high schools	• 204 sprains at baseline • 198 sprains in analysis	High-school athletes	≤ 24 hours	Time to return to play: same day (23.7%), next day (21.2%), 3 days (29.3%), 7 days (11.6%), 10 days (8.6%) or > 22 days (5.6%)	Time to return to play: same day, next day, 3 days, 7 days, 10 days, 21 days or > 22 days

ATFL, anterior talofibular ligament; FAD, first aid department; NR, not reported; NWB, non-weight bearing.

Risk-of-bias assessment

Table 3 shows the outcome of the risk-of-bias assessments. One study was judged as being at a low risk of bias,[15] five were judged as being at a moderate risk of bias,[30,32,34,36,37] and three studies were judged as being at a high risk of bias.[31,33,35] Incomplete and/or inadequate reporting standards were common issues; for example, it was difficult to identify whether prognostic factors were eliminated because of statistical reasons or poor clinical utility. No studies reported on the performance of the prognostic models by using methods to assess internal or external validation.

Prognostic factors identified

Prognostic factors included in the final models for each included study are shown in *Tables 4* (short term), *5* (medium term) and *6* (long term).

Prognostic factors for short-term recovery (≤ 8 weeks)

Five studies investigated prognostic factors for short-term recovery (*Table 4*).[30–32,35,36]

de Bie *et al.*[30] reported that having a baseline Ankle Function Score (AFS) of ≤ 35 points was a prognostic factor for non-recovery at 2 weeks. A combination of an AFS of ≤ 35 points, higher severity grading by a doctor and a higher palpation/ligament stress test score was included in the final model for the 4-week time point. van der Wees *et al.*[35] reported that a baseline AFS of ≤ 40 points was a prognostic factor for non-recovery at 2 weeks. Wilson and Gansneder[31] reported that greater range-of-motion loss and a greater extent of swelling were prognostic factors for a longer duration of disability. They also reported

TABLE 3 Risk-of-bias assessment of the nine included studies, in accordance with the QUIPS tool

Study	Study participation	Study attrition	Prognostic factor measurement	Outcome measurement	Study confounding	Statistical analysis and reporting	Overall risk of bias
de Bie *et al.*[30]							
Wilson and Gansneder[31]							
Cross *et al.*[32]							
Akacha *et al.*[15]							
Langner *et al.*[33]							
van Middelkoop *et al.*[34]							
van der Wees *et al.*[35]							
O'Connor *et al.*[36]							
Medina McKeon *et al.*[37]							

Note
Green, low risk of bias; yellow, moderate risk of bias; red, high risk of bias.

TABLE 4 Prognostic factors for short-term (≤ 8 weeks) outcome in acute lateral ankle sprain

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with short-term outcome
de Bie *et al.*[30]	Healed or not healed at 2 and 4 weeks. Healed = AFS of > 75 points (on a scale of 0–100 points) and palpation/ligament stress test score of < 2 (0–12)	AFS (0–100) of ≤ 35 points; doctor severity grading (0–10); palpation/ ligament stress test score (0–12)	Multivariable logistic regression	2 weeks: baseline AFS of ≤ 35 points predicted poor outcome ('not healed'). Sensitivity = 97%, specificity = 100% 4 weeks: combined baseline AFS of ≤ 35 points, severity grading and palpation/ligament stress test score predicted poor outcome ('not healed'). Sensitivity = 81%, specificity = 80%
Wilson and Gansneder[31]	Number of days to return to full sports practice or competition [11.9 days (SD 6.6 days)]	Joint swelling (ml), sagittal plane ROM loss (degrees), objective WB activity score (0–6), self-reported athletic ability score (VAS, 0–100 points)	Hierarchical regression	Combined swelling (β = –0.02) and ROM loss (β = –0.08). $R^2 = 0.34$; $p = 0.023$ Combined WB activity score (β = –0.55) and self-reported ability score (β = –0.39). $R^2 = 0.33$; $p = 0.004$ Combined swelling, ROM loss, WB activity score and self-reported athletic ability score. $R^2 = 0.59$; $p = 0.001$
Cross *et al.*[32]	Number of days to return to sport [14.7 days (SD 8.8 days)]	SF-36 PF (0–100), self-reported global function (0%–100%), objective ambulation status (1–7)	Univariate regression, stepwise multivariable regression	SF-36 PF: $R^2 = 0.28$; $p = 0.016$. Self-reported global function: $R^2 = 0.22$; $p = 0.036$ Objective ambulation status: $R^2 = 0.22$; $p = 0.019$ Combined SF-36 PF, self-reported global function and objective ambulation status: $R^2 = 0.34$; $p < 0.01$
van der Wees *et al.*[35]	Global perceived effect of ≥ 2 (1 = recovered, 2–7 = not recovered) at 2 weeks	AFS (0–100) of ≤ 40 points	Sensitivity and specificity	2 weeks: baseline AFS of ≤ 40 points predicted recovery status. Sensitivity = 76%, specificity = 63%
O'Connor *et al.*[36]	Karlsson function score (0–100) at 4 weeks	Age (years), injury grade (1, 2, 2+), WB status (FWB, FWB with pain, PWB, NWB)	Univariate regression, stepwise multivariable regression	4 weeks: combined age (β = –0.32; $p = 0.001$), injury grade (β = –0.23; $p = 0.003$) and WB status (β = –0.34; $p = 0.038$). $R^2 = 0.34$; $p < 0.01$

FWB, full weight bearing; NWB, non-weight bearing; PWB, partial weight bearing; ROM, range of motion; SF-36 PF, Short Form questionnaire-36 items, physical function scale; VAS, visual analogue scale; WB, weight-bearing.

greater functional limitations, measured on an objective six-item weight-bearing activity score and on a self-reported current athletic ability rating, as a prognostic factor.[31] The effect of these ankle impairment and functional limitation prognostic factors was additive, and together they explained 59% of the variance in disability duration.[31] Cross *et al.*[32] reported the baseline prognostic factors of lower self-reported physical function, self-reported global function and objectively measured ambulation status as being associated with a greater number of days to return to sport.

O'Connor *et al.*[36] reported that baseline prognostic factors of greater age, more severe injury grade and poorer weight-bearing status were associated with lower subjective ankle function at 4 weeks post injury.

Prognostic factors for medium-term recovery (≤ 4 months)

O'Connor *et al.*[36] reported that greater age, poorer weight-bearing status and non-inversion injury mechanism were prognostic factors for poorer subjective function at 4 months' follow-up (*Table 5*). They also identified medial joint line pain on palpation and pain on weight bearing during ankle dorsiflexion at 4 weeks as prognostic factors for poorer subjective function at 4 months.[36]

Prognostic factors for long-term recovery (> 4 months)

Three studies[15,33,34] reported prognostic factors for long-term recovery (*Table 6*). Akacha *et al.*[15] demonstrated that higher age and female sex were prognostic factors for slower and incomplete recovery. Langner *et al.*[33] reported that more severe grade of injury, greater number of injured ligaments and presence of a bone bruise [all determined with magnetic resonance imaging (MRI)] were associated with greater time taken to return to sports activities. van Middelkoop *et al.*[34] reported that none of the candidate prognostic factors measured at baseline was associated with outcome at 12 months' follow-up.

Table 7 is on overview of all the prognostic factors investigated and the time points at which they were assessed, and indicates if the methods used within the study did or did not find evidence of an association between the variable and the outcome.

TABLE 5 Prognostic factors for medium-term (≤ 4 months) outcome in acute lateral ankle sprain

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with medium-term outcome
O'Connor *et al.*[36]	Karlsson AFS (0–100) at 4 months	• Baseline: age (years), WB status (FWB, FWB with pain, PWB, NWB); injury mechanism (inversion/other) • 4 weeks: pain on WB during ankle dorsiflexion; medial joint line pain (yes/no)	Univariate regression, step-wise multivariable regression	• 4 months: baseline combined age ($\beta = -0.26$; $p = 0.01$), WB status ($\beta = -0.23$; $p = 0.25$) and injury mechanism ($\beta = -0.25$; $p = 0.17$). Adjusted $R^2 = 0.34$; $p < 0.01$ • 4 months: 4-week combined pain on WB during ankle dorsiflexion ($\beta = 0.60$; $p < 0.001$), medial joint line pain ($\beta = 0.24$; $p = 0.07$). Adjusted $R^2 = 0.49$; $p < 0.01$

FWB, full weight bearing; NWB, non-weight bearing; PWB, partial weight bearing; WB, weight-bearing.

TABLE 6 Prognostic factors for long-term (> 4 months) outcome in acute lateral ankle sprain

Study	Primary outcome measure	Variables in final model	Analysis	Prognostic factors in final models associated with long-term outcome
Akacha *et al.*[15]	FAOS symptoms subscale (0–100: 0 = extreme symptoms, 100 = no symptoms)	Age, sex	Non-linear mixed model	• Greater age and female sex associated with slower and incomplete recovery • Greater age (β = –0.01, 95% CI –0.12 to –0.004) Female (β = –0.06, 95% CI –0.01 to –0.002)
Langner *et al.*[33]	Time to return to sports activities	MRI-grading of ligamentous injury (1–3: 1 = stretching, 2 = partial tear, 3 = complete tear); number of injured ligaments; presence of bone bruise	Multivariable regression	• MRI-grading of ligamentous injury, $R^2 = 0.45$; $p < 0.01$ • Number of injured ligaments, $R^2 = 0.35$; $p < 0.01$ • Bone bruise, $R^2 = 0.32$; $p < 0.01$
van Middelkoop *et al.*[34]	Self-reported recovery (NRS, 0–10: 0 = not recovered; 10 = completely recovered) at 12 months	Re-sprain within 3 months; pain at rest at 3 months (NRS, 0–10)	Multivariable regression	• 12 months: re-sprain within 3 months (β = –1.64, 95% CI –3.11 to –0.16) • Pain at rest at 3 months (β = –0.69, 95% CI –1.08 to –0.29)

CI, confidence interval; NRS, numerical rating scale.

TABLE 7 Summary of all formally investigated prognostic factors across the included studies

Prognostic factor assessed	Study								
	de Bie *et al.*[30]	Wilson and Gansneder[31]	Cross *et al.*[32]	Akacha *et al.*[15]	Langner *et al.*[33]	van Middelkoop *et al.*[34]	van der Wees *et al.*[35]	O'Connor *et al.*[36]	Medina McKeon *et al.*[37]
Age				LT ✓		LT✗		ST ✓	
								MT ✓	
AFS	ST ✓					LT✗		✗	
Active ROM for injured leg			ST✗						
Active ROM for uninjured leg			ST✗						
BMI						LT✗		ST ✗	
Clinical severity grading	ST ✓							ST ✓	
Dorsiflexion muscle strength for injured leg			ST✗						
Dorsiflexion muscle strength for uninjured leg			ST ✗						
Gait pattern						LT✗			
Sex				LT ✓		LT✗		✗	✗
Global function question			ST ✓						
GPE							ST ✓		
Injury grade						LT✗		ST ✓	
Instability						LT✗			
Mechanism of injury								MT ✓	
Medial joint line pain on palpation								MT ✓	
Olerud–Molander[38] Ankle Score							ST ✓		
MRI grading of bone bruise					LT✓				
MRI grading of number of injured ligaments					LT✓				
MRI severity grading of ligamentous injury					LT✓				

continued

TABLE 7 Summary of all formally investigated prognostic factors across the included studies (*continued*)

Prognostic factor assessed	Study								
	de Bie *et al.*[30]	Wilson and Gansneder[31]	Cross *et al.*[32]	Akacha *et al.*[15]	Langner *et al.*[33]	van Middelkoop *et al.*[34]	van der Wees *et al.*[35]	O'Connor *et al.*[36]	Medina McKeon *et al.*[37]
Pain at rest						LT✗			
Pain at rest at 3 months						LT✓			
Pain on weight-bearing ankle dorsiflexion								MT ✓	
Pain while running						LT✗			
Pain while walking						LT✗			
Palpation score	ST ✓								
Patient-specific complaints							ST ✓		
Plantar flexion muscle strength for involved leg			ST ✗						
Plantar flexion muscle strength for uninvolved leg			ST ✗						
Previous ankle sprain history								✗	ST ✗
Reduced ROM		ST ✓							
Referrals									ST✗
Re-sprain within 3 months						LT✓			
Return to full sports activities					LT✗				
Return to work on full duties					LT✗				
Self-reported global function							ST ✓		
Self-reported athletic ability		ST ✓							
Self-reported physical limitations			ST ✓						
Setting						LT✗			
SF-36 Physical Function Scale			ST ✗						
Side-hop test								✗	

Prognostic factor assessed	Study								
	de Bie *et al.*[30]	Wilson and Gansneder[31]	Cross *et al.*[32]	Akacha *et al.*[15]	Langner *et al.*[33]	van Middelkoop *et al.*[34]	van der Wees *et al.*[35]	O'Connor *et al.*[36]	Medina McKeon *et al.*[37]
Sport load						LT✗			ST✗
Subjective recovery						LT✗			
Swelling		ST ✓				LT✗			
Treatment/randomisation group				LT ✗		LT✗			
VAS for pain			ST ✗						
Activity score		ST ✓							
Weight-bearing status	ST ✓		ST ✓					ST ✓	
								MT ✓	
Work load						LT✗			

BMI, body mass index; GPE, global perceived effect; LT, long term; MT, medium term; ROM, range of motion; SF-36, Short Form questionnaire-36 items; ST, short term; VAS, visual analogue scale.

Note

Dark green, low risk of bias; light green, moderate risk of bias; blue, high risk of bias.

✓ Prognostic factor assessed and included in final models, with evidence of statistical association with outcome.

✗ Prognostic factor assessed but not found to be statistically associated with outcome (typically dropped in univariable analysis before multivariable modelling, or dropped during multivariable modelling analysis).

Discussion

Across the included studies, a wide range of prognostic factors was investigated. The prognostic factors that were analysed varied considerably between studies, with no common framing across the studies. Owing to the methodological issues identified in the majority of included studies, it is important that the evidence of statistical associations between the candidate prognostic factors and the outcomes reported should be interpreted with caution.

Age was identified as an independent prognostic factor in one study rated as having a low risk of bias[15] and in another study[36] rated as being at a moderate risk of bias. Higher baseline age was associated with poor recovery at short-,[36] medium-[36] and long-term follow-up.[15] Injury severity was reported as a prognostic factor in two studies by clinical symptoms[30,36], but in another study[33] MRI was used to grade severity. Clinical assessments may be subjective to some extent, but sensitive investigations, such as MRI, are not readily available in acute settings. Furthermore, the insufficient evidence for diagnostic imaging findings as prognostic factors highlights that structural pathology may not be indicative of clinical severity. A lack of association between structural changes in the ankle and persistent ankle impairments has been reported.[39]

Measures obtained somewhat later after injury (4 weeks for predicting outcome at $\leq$ 4 months;[36] 3 months for predicting outcome at 12 months[34]) appeared to have better prognostic value than in the early acute stage, indicating that the timing of the measurement can influence the value of prognostic factors. The challenge of using measures taken later after injury is that this could delay decisions about monitoring and early intervention.

Limitations of ankle sprain prognostic factor studies

In the majority of the included studies, follow-up was only short term, and was discontinued at a time when symptoms were still prominent and resolving, and hence recovery was quite variable. Methodological shortcomings were evident across the studies, for example, none reported an assessment of interval validity or attempted an external validation of its models. Adjustments for confounding factors such as time since injury, were not employed. Regression analyses were often not reported in sufficient detail to identify whether prognostic factors were eliminated because of small sample size or poor clinical utility. Two studies[30,35] dichotomised a continuous outcome measure. The cut-off points that were used were not well justified or prespecified.

The study[15] judged as being of high quality tended to report conservative estimates of associations between predictors and outcome. However, a limited range of prognostic factors was investigated.

Although a wide range of prognostic factors have been investigated, the limitations of previous studies highlight the need for large-scale studies that employ robust prognostic research methods[10] and adhere to recognised reporting guidelines.[25] The systematic review that we conducted did provide some evidence to inform the decision making processes within the consensus exercise.

Chapter 4 Consensus meeting

Introduction

In this chapter, we report the findings of a UK-based consensus meeting that assisted in determining which prognostic factors should be considered as candidates in the SPRAINED prognostic model. There is no universally accepted method on how best to develop a prognostic model.[40] Current recommendations for this include using variables that have already demonstrated prognostic value (see *Chapter 3*) and including other clinically plausible variables.[41] Therefore, our aim was to use a triangulation of methods to ensure that a comprehensive selection of prognostic factors was considered for inclusion in the SPRAINED prognostic model. First, we used the results from preliminary analyses of data from a previous large-scale clinical trial[5] involving people with acute lateral ankle sprains attending EDs to explore which prognostic factors could be important for predicting recovery at 9 months after injury (see *Chapter 5* for details). Second, we used the results of our systematic literature review of studies, investigating prognostic factors for recovery (see *Chapter 3*) to elucidate which prognostic factors had been previously identified, and the level of evidence for these factors. Third, we used a consensus meeting to triangulate these factors with clinical and patient/public opinion.

In order to optimise the development of the SPRAINED prognostic model, we aimed to obtain interpretations of these sources of evidence from a range of key stakeholders and achieve consensus on which baseline and delayed prognostic factors should be included in the prognostic model that was to be evaluated in the external validation study (see *Chapter 6*).

Methods

A variety of methodologies for achieving consensus exist (e.g. Delphi methods, discrete choice experiments and face-to-face methods), but there is no agreed optimum approach on how to synthesise judgements when a state of uncertainty exists.[42] We chose to use a mNGT because it provided a structured scientific process, which incorporated the private views of individual participants, and facilitated discussion leading to an aggregated group judgement. The mNGT was originally reported by Delbecq *et al.*[43] and has since been refined and utilised in a range of musculoskeletal research settings, most notably in the Outcome MEasures in Rheumatoid Arthritis (Rheumatology) Clinical Trials (OMERACT) initiative.[44] In mNGT, individual participants express views via a questionnaire before a face-to-face meeting, in which findings are fed back, structured discussion is facilitated and then a final vote is taken of individual views.[45]

Participants

We aimed to recruit a range of key stakeholders, including patient and public representatives, health-care professionals and clinical researchers, to represent a range of parties involved in ankle sprain care and research in the UK NHS. We invited 30 individuals to participate, including a variety of health-care professionals from across the UK who worked in ambulance services, general practice, radiology, emergency and trauma surgery departments, as well as clinical researchers. We also aimed to recruit patient and public representatives from the south central area of the UK who had experience of an ankle sprain or were able to represent an individual or group that had such experience. We placed adverts for patient and public representatives in local supermarkets, in the John Radcliffe Hospital, on the People in Research[21] website and the NIHR Research Design Service South Central's mailing list.

Facilitators

The SPRAINED study team facilitators were guided by a lead facilitator (KH) with experience in conducting mNGT processes in musculoskeletal research.[46] Additional facilitators were provided with a standardised brief to follow during the meeting and supervised by the lead facilitator.

Consensus process

We conducted the consensus process in three main stages, outlined in the following sections.

Preparation and supply of information

Participants were provided with an electronic information pack 10 days before a face-to-face meeting. This pack consisted of a summary of the SPRAINED study to date, findings from the systematic review of prognostic factors for acute ankle sprains (see *Chapter 3*), preliminary findings from statistical modelling of the CAST data set (see *Chapter 5*) and a pre-meeting questionnaire.

Completion of pre-meeting questionnaire

The pre-meeting questionnaire was developed with two key sections (see *Appendix 3*). The first section elicited the participants' opinions on which prognostic factors were important for recovery following acute ankle sprain. Data from the systematic review and statistical modelling were utilised to generate a list of 14 predefined factors. Participants were also given the facility to nominate unlisted factors. Response options were provided in the form of the 9-point Grading of Recommendations Assessment, Development and Evaluation (GRADE)[47] scale (1 to 3, not important; 4 to 6, important but not critical; 7 to 9, critical) with importance defined as 'How important do you think [prognostic variable] is a factor in recovering from an ankle sprain?' A 'don't know' response box was also provided as an option.

A second section was developed to enquire when and how additional delayed information should be obtained. This was informed by studies included in the systematic review (see *Chapter 3*) that demonstrated that information collected after baseline improved prognostic model accuracy. The questions were in the form 'If we were to collect further information like this, how many weeks after the initial visit do you think we should collect this information?' (response options ranged from '1 week' to '6 weeks') and 'How should we collect this information?' (response options were hospital visit, postal questionnaire, online questionnaire, telephone questionnaire). The pre-meeting questionnaire was piloted with two potential participants (one patient representative and one clinical researcher), who provided comment on structure, content and clarity.

The consensus process participants were asked to complete and return the questionnaire in electronic form before the meeting. Data were analysed before the meeting to summarise the distribution of ratings for each prognostic factor, including the group median and interquartile range. The importance of a factor was deemed to be 'critical' if the group median score ranged between 7 and 9.[48]

Consensus meeting

This was a 1-day meeting, held in Oxford, UK. The meeting had three sections.[49] At the start of the meeting a detailed explanation of the systematic review and preliminary statistical modelling was provided, followed by a summary of responses to the pre-meeting questionnaire (participants were also provided with copies of their own individual responses).

The second section consisted of two rounds of structured facilitator-led discussions that aimed to identify the most important prognostic factors measured initially, and which delayed prognostic factors should be collected and how. The participants were divided into three groups (to which participants were pre-assigned to ensure a mixture of clinicians, researchers and patient representatives) and were asked to rank a maximum of 10 important prognostic factors (from the 14 factors identified from the pre-meeting questionnaire) and five important additional prognostic factors (from the 20 nominated in the pre-meeting questionnaire). Ten points and five points, respectively, were awarded to the most important factor, and one to the least important factor. Each round of group discussions was immediately followed by a plenary session to feed back results of the group discussions to the entire group.

Finally, a session was convened during which a final voting process was undertaken: each participant indicated whether or not each factor should be included in the prognostic model. The number of votes allowed was limited to 10 per individual. This was completed independently on paper questionnaires and then collated. Factors with ≥ 70% agreement across participants were considered as critically important to consider in the validation study.[50]

Results

Participants

Of the 30 individuals invited, 25 clinicians and clinical researchers agreed to participate comprising: paramedics ($n = 6$), physiotherapists ($n = 6$), ED nurses ($n = 4$), ED consultants ($n = 5$), radiology consultant ($n = 1$), trauma and orthopaedic consultant ($n = 1$) and clinical researchers ($n = 2$). Three patient and public representatives responded to the advertisements, but only one was able to attend the consensus meeting. The pre-meeting electronic questionnaire was returned by 17 individuals; and 18 individuals attended the meeting and participated in the first two rounds of group discussions. Two participants were unable to complete the final round of individual voting. Hence, only 16 participants voted for the factors that had been prioritised throughout the day.

Pre-meeting questionnaire results

The results of the electronic pre-meeting questionnaire are shown in *Table 8*. Three baseline factors were rated as critically important (scoring between 7 and 9) and the remainder as important but not critical (scoring between 4 and 6). The respondents nominated 20 additional factors that were deemed critically

TABLE 8 Findings from the pre-meeting questionnaire including ratings of importance for baseline prognostic factors and additional nominated factors

Question	Prognostic factor[a]	Median (IQR)	Minimum, maximum
1	Time between injury and presenting to ED	5 (4, 6)	1, 7
2	Pain severity	5 (4, 6)	2, 7
3	Pain on weight bearing	7 (4, 7)[b]	2, 8
4	Weight-bearing status in ED	6 (5, 7)	2, 9
5	Amount of ankle movement (dorsiflexion)	4.5 (3, 6)	2, 7
6	Amount of ankle movement (plantarflexion)	5 (3, 6)	2, 8
7	Abnormal imaging findings	6 (5, 8)	3, 9
8	Age	6 (5, 8)	2, 9
9	BMI	7 (5, 7)[b]	2, 8
10	Working status	5 (4, 6)	2, 9
11	Level of education	4 (3, 5)	1, 7
12	Mechanism of injury	6 (4, 7)	2, 8
13	Repeatedly sprained ankle previously	7 (5, 8)[b]	5, 9
14	Reporting of catching or locking of the ankle	5.5 (5, 6)	3, 7

BMI, body mass index; IQR, interquartile range.

a Other factors nominated in the pre-meeting questionnaire were history of chronic pain ($n = 2$), comorbidities (including osteoporosis) ($n = 3$), sporting participation, swelling ($n = 2$), anterior talofibular ligament vs. posterior talofibular ligament injury, weight-bearing status immediately post injury, occult fracture, other soft tissue damage, syndesmotic sprain, anxiety ($n = 3$), perception of injury severity, recovery expectations ($n = 2$), desire to get better, self-efficacy, beliefs, coping styles, ability to exercise despite pain, requiring regular analgesia, physiotherapy/rehabilitation referral ($n = 3$), adherence to advice.

b Rating critically important.

important. There was a varied response to when and how delayed prognostic factors should be collected. The most frequent preferences were 4 weeks post injury and by telephone.

Consensus meeting results

Eighteen participants, divided into three groups, participated in the two rounds of facilitated discussions and prioritisation exercises. Some groups were unable to agree on or did not use the maximum number of ranks. Priority rankings of the prognostic factors rated by the three groups of key stakeholders are shown in *Table 9*. The prognostic factors of the highest priority included repeatedly spraining ankle previously, older age and mechanism of injury. Only 6 of the 20 additional factors nominated in the pre-meeting questionnaire were deemed high priority for inclusion in the prognostic model: (1) occult fracture/diagnostic imaging result, (2) history of chronic pain/problems, (3) desire to get better, (4) psychosocial factors about recovery, (5) weight-bearing status immediately post injury and (6) self-efficacy. Following the facilitated discussions, 16 participants completed the final vote for which factors to include in the

TABLE 9 Results of final voting for prognostic factors. Dichotomous responses (yes/no)

	Results of meeting			
	Section 2 – priority rank (10 highest, 1 lowest)			Section 3 – votes for inclusion of factor in prognostic model, *n* (%)
Prognostic factor	Group 1	Group 2	Group 3	
Age	1	10	8	16 (100)
BMI	7	–	7	16 (100)
Repeatedly sprained ankle previously	8	9	10	16 (100)
Weight-bearing status in ED	–	8	6	16 (100)
Mechanism of injury	6	–	9	14 (88)
Pain on weight bearing	10	–	4	14 (88)
Working status	5	6	–	14 (88)
Pain severity	–	–	3	13 (81)
Time between injury and presenting to ED	–	7	2	7 (44)
Amount of ankle movement (dorsiflexion)	–	–	5	7 (44)
Abnormal imaging findings	9	–	–	7 (44)
Amount of ankle movement (plantarflexion)	–	–	–	4 (25)
Level of education	–	–	–	2 (13)
Reporting of catching or locking of the ankle	–	–	–	0 (0)
Additional prognostic factors nominated in pre-meeting questionnaire	***Group 1 (5 highest, 1 lowest)***	***Group 2 (5 highest, 1 lowest)***	***Group 3 (5 highest, 1 lowest)***	
Psychosocial factors about recovery	2	5	–	12 (75)
Occult fracture/diagnostic imaging result	5	–	–	
History of chronic pain/problems	4	3	–	
Desire to get better	3	–	–	
Weight-bearing status immediately post injury	1	–	–	
Self-efficacy	–	4	–	

Note
Shading indicates factors with ≥ 70% agreement across participants (considered critically important).

prognostic model. Participants agreed to include 8 of the 14 originally proposed prognostic factors in the prognostic model: (1) pain intensity, (2) pain intensity on weight bearing, (3) weight-bearing status in the ED, (4) age, (5) body mass index (BMI), (6) working status, (7) mechanism of injury and (8) repeatedly sprained ankle previously. Only one additional factor nominated from the pre-meeting questionnaire was agreed on for inclusion in the prognostic model – psychosocial recovery factors (see *Table 9*). No delayed factors were agreed on for inclusion in the prognostic model.

Discussion

This chapter described the consensus-based approach employed in the development of the SPRAINED prognostic model. We identified eight baseline factors that were deemed critical for the identification of people likely to have a poor recovery. These factors span pre-injury, sociodemographic, psychosocial and clinical assessment factors, encompassing a holistic biopsychosocial model of recovery.[51]

Only one prognostic variable not included in the CAST data set (see *Chapter 5*) was deemed important enough to be added to the prognostic variables collected in the external validation study (see *Chapter 6*) to enable a later investigation into this prognostic factor. It was agreed that participants should be asked how long they expected to take to recover from their ankle sprain, which aimed to capture the person's psychological state and perceptions in the acute phase. No additional delayed factors were rated as being critical for inclusion in the model.

The results of our meeting were strengthened by the use of a diverse group of clinical and research practitioners, in addition to a patient and public representative. We also had the opportunity to test the structure and content of the questions that we presented to the group for voting. The limitations of our approach include the lower than anticipated number of patient participants with direct experience of short- or long-term limitations attributable to an ankle sprain. This may have provided a broader perspective relevant to this patient population. A limitation of the mNGT is the short time constraints, limiting the reiterations of the discussion process and time that participants have to reflect and achieve consensus. The pragmatic approach used may have influenced the length of the group discussions and, consequently, the final results.

The findings of this consensus meeting were used in combination with the findings of the systematic review (see *Chapter 3*) and the statistical analysis development (see *Chapter 5*) to inform which additional factors could be included in the model assessed during the external validation study (see *Chapter 6*). The main impact of the meeting was a strengthening of the evidence regarding prognostic factors already considered candidates for the model and, importantly, the addition of a question to consider the psychosocial status around the expectation of recovery, as a reflection of wider beliefs and anxieties about the injury and recovery.

The size of the CAST data set was known ahead of all the modelling processes; this fact allowed us to prespecify, with the use of simple rules, the number of variables that could plausibly be considered as candidates in the internal validation. The consensus exercise was essential in determining the priority variables to consider, and the acceptability and method of testing the variable, from both the clinical and patient community perspectives. There were a few exceptions to this process. The research team considered that it was necessary to include commonly used clinical examination procedures during the consensus stage. Ultimately, neither the systematic review nor consensus meeting identified these as important. The patchiness and limited scope of existing evidence and relatively limited sampling for the consensus group meant that the possibility of falsely excluding variables might be high; therefore, we erred on the side of caution.

Chapter 5 Development and internal validation of the SPRAINED prognostic models in the CAST data set

Introduction

This chapter describes the development and internal validation of the two prognostic models to identify people at risk of poor outcome after an acute ankle sprain. The development of the two models followed the same steps using the same data set, and considered the same candidate predictors, but had different definitions of outcome. Data from CAST, a RCT on the effectiveness of three different mechanical supports compared with a double-layer tubular compression bandage for the initial management of severe ankle sprains, were used to develop both models.[18]

The initial selection of variables for testing in the CAST data (before and for the consensus review) was guided by the systematic review (see *Chapter 3*) and analysis of the data set. The final selection of variables for testing in internal validation was informed by the results of the consensus meeting (see *Chapter 4*).

Methods

Individual participant data used to develop the models (study population)

CAST was a pragmatic, multicentre RCT, with blinded assessment of the outcome, designed to estimate the clinical effectiveness and cost-effectiveness of three different types of mechanical ankle support [Aircast® ankle brace (DJO Incorporated, Vista, CA, USA), Bledsoe® boot (Bledsoe Boot Systems, Grand Prairie, TX, USA) or 10-day below-knee cast] in the treatment of severe ankle sprain (defined as an injury of grade 2 or 3, without fracture) compared with a double-layer tubular compression bandage.

The trial population comprised 584 individuals aged $\geq$ 16 years attending EDs in the UK with an ankle sprain and an inability to fully bear weight on the injured ankle at the time of presentation to the ED and their review clinic appointment (the trial's baseline assessment). People were excluded if they presented with an ankle fracture (apart from flake fractures of $\leq$ 2 mm), any other recent fracture, any contraindication to any of the four arms of the trial, poor skin viability preventing splinting or casting, or if their injury occurred > 7 days before the first presentation at the recruiting ED.

The different time points in CAST and a summary of the data collected at each point are defined in *Table 10*.

Definition of the primary outcomes

Ankle function at 9 months after ankle sprain was the primary outcome for CAST. For the SPRAINED study, our primary outcome was 'poor outcome'. We used two definitions of poor outcome that were based on key indicators of poor function and instability of the joint, which is typified by recurrent sprains or a significant lack of confidence in the ankle (a persistent feeling of giving way), with or without chronic pain. The selection of these outcome indicators is supported by evidence from van Rijn *et al.*,[52] who reported that recovery was most closely associated with improvements in pain and giving way, and Wikstrom *et al.*,[7] according to whom pain and instability are of greatest concern to patients. The definitions were considered and agreed by the patient and public involvement group convened for the SPRAINED study.

Data to classify these outcomes were collected in the CAST data set as outlined in the following sections.

TABLE 10 Definitions of time points in CAST

Time point	Definition	Information collected
1. First contact with participants (ED presentation)	Individuals with an ankle sprain attending an ED that was recruiting for the trial were assessed for eligibility by medical staff, who also completed a standard pro forma with some basic clinical and sociodemographic information. Information on the trial and an invitation to join the study was given to eligible individuals together with the participant information leaflet	Initial eligibility criteria check (people aged ≥ 16 years, attending EDs no more than 7 days after injury, with sprain – not fracture – of the ankle and unable to fully bear weight at presentation); clinical examination and injury-related information; and sociodemographic data
2. Follow-up clinic at 2 or 3 days after ED attendance (baseline assessment)	Final eligibility check and informed consent obtained from those willing to enter the trial. Short interview performed by the research physiotherapist to ensure eligibility and, after randomisation, participants completed a baseline questionnaire. The interventions were applied in the ED by an appropriately trained health professional after baseline data collection and randomisation	Data on the main candidate predictors for the prognostic model, including age, sex, height, weight, ethnicity, pre-injury quality of life, mobility, engagement in sports activities, usual occupation and employment. Data on injury presentation, indicators of current mobility levels, pain, and weight-bearing status were also collected
3. Outcome measurements (follow-up assessments)	All outcome measurements were taken at 4 weeks, 12 weeks and 9 months	• **Primary outcomes:** FAOS; FLP • **Secondary outcomes:** The SF-12 scale for health-related quality of life, EQ-5D for economic evaluation, VAS to estimate the pain at rest and when bearing weight, the Benefit Scale to rate the benefit received from the treatment, and health-care resource use

EQ-5D, EuroQol-5 Dimensions; FLP, Functional Limitation Profile; SF-12, Short Form questionnaire-12 items; VAS, visual analogue scale.

Severe persistent pain

Severe persistent pain was defined on the basis of the response given to the question 'How often do you experience foot/ankle pain?' from the FAOS.[19] The five available response options to this question were (1) never, (2) monthly, (3) weekly, (4) daily or (5) always. Participants who answered 'daily' or 'always' were considered to have severe persistent ankle pain.

Severe functional difficulty

Severe functional difficulty was defined on the basis of the response given to the question 'In general, how much difficulty do you have with your foot/ankle?' from the FAOS.[19] The five available response options to this question were (1) none, (2) mild, (3) moderate, (4) severe or (5) extreme. Participants who answered 'severely' or 'extremely' were considered to have severe functional difficulty with the ankle.

Significant lack of confidence

Significant lack of confidence was defined on the basis of the response given to the question 'How much are you troubled with lack of confidence in your foot/ankle?' from the FAOS.[19] The five available responses to this question were (1) not at all, (2) mildly, (3) moderately, (4) severely or (5) extremely. Participants who answered 'severely' or 'extremely' were considered to have a significant lack of confidence in the ankle.

Recurrence of injury

Recurrence of injury was defined as a new injury of the same nature (acute ankle sprain) to the same ankle, occurring after the initial assessment (baseline) and up to 9 months after the date of the first injury. Data on this event were collected by asking a specific question: 'Have you had another injury to the same ankle?'.

Composite outcome generation

Two different composite outcomes were generated, focusing on self-reported recovery (outcome 1), and self-reported recovery plus whether or not participants had experienced a recurrence of their ankle sprain during the 9-month follow-up period (outcome 2). The investigation of these two different composite outcomes was conducted because recurrence of sprain was considered a sufficiently different clinical issue that could potentially widen the range of patients considered as having a poor outcome, and therefore warranted consideration separately.

Outcome 1

The first model was developed to predict a composite outcome (hereafter referred to as outcome 1) representing the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence. First, individual binary outcomes (yes or no) were generated to indicate the presence of each symptom, in accordance with the criteria described earlier in this section. A single composite binary outcome (outcome 1) was then created to indicate the presence (yes or no) of one or more of these symptoms.

Outcome 2

The second model was developed to predict a composite outcome (hereafter referred to as outcome 2) representing the presence of at least one of the following symptoms or clinical events at 9 months after injury: persistent pain, functional difficulty, lack of confidence or recurrence of injury. First, individual binary outcomes (yes or no) were generated to indicate the presence of each symptom or clinical event, in accordance with the criteria described earlier in this section. A single composite binary outcome (outcome 2) was then created to indicate the presence (yes or no) of one or more of these symptoms or events.

The proportion of these outcomes observed in the CAST data set for both outcomes 1 and 2 and the number of symptoms at 9 months after injury are described in *Table 11*.

Available candidate predictors and initial selection of variables for modelling

A complete list of the 16 available variables in the ED pro forma is provided in *Box 1*, and a complete list of the 154 available variables in the CAST baseline data set is provided in *Box 2*. Variables available in the CAST data set included sociodemographic indicators (age, sex, BMI, education, employment status); pre-injury quality of life, mobility and lifestyle (e.g. engagement in sports activities); clinical data on injury presentation; and indicators of current mobility levels, pain and weight-bearing status. From these lists,

TABLE 11 Proportion of outcomes (and components) observed in the CAST data set

	Outcomes observed, *n* (%)						
Symptoms/events	None present	1 present	2 present	3 present	4 present	Any present	Missing, *n* (%)
Outcome 1 (pain, lack of confidence or general difficulty)	324 (55.48)	68 (11.64)	19 (3.25)	29 (4.97)	–	116 (19.86)	144 (24.7)
Outcome 2 (pain, lack of confidence, general difficulty or re-injury)	300 (51.37)	82 (14.04)	26 (4.45)	23 (3.94)	9 (1.54)	140 (23.97)	144 (24.7)

BOX 1 List of candidate predictor variables from the ED presentation data set

Emergency department pro forma

1. Date of birth.
2. Sex (male/female/no response).
3. Date of ED visit.
4. Date of injury.
5. Location of pain.
6. Anterior drawer test (positive/painful/negative/no response).
7. Talar tilt test (positive/painful/negative/no response).
8. Tenderness of proximal fibular (positive/painful/negative/no response).
9. Weight-bearing ability (full/partial/none/no response).
10. Radiograph (yes/no/no response).
11. Crutches (yes/no/no response).
12. Reason for not entering trial (ankle fracture/other recent fracture/contraindication to intervention/poor skin viability/> 7 days from injury to assessment/other).
13. Additional information (if other).
14. Recruiting centre.
15. Date of trial clinic.
16. Days from injury to assessment.

BOX 2 List of candidate predictor variables from the baseline assessment data sets

Identifier variables

1. Trial centre.
2. Patient's identification.
3. Date of assessment.
4. Randomisation group.
5. Treatment received.
6. Calendar code.
7. Calendar colour.
8. Indicator of pilot study phase (I/II/main trial).
9. Response at baseline (yes/no).

Background information form

10. Age (years).
11. Sex (male/female).
12. Ethnic group (white/black-caribbean/black-African/black-other/Indian/Pakistani/Bangladeshi/Chinese/other).
13. Ethnic group details (if other).
14. First language (English/other European/Gujarati/Hindi/Punjabi/Urdu/Bengali/other).
15. First language additional information (if other).
16. Able to answer English questions (yes/no).
17. Current employment status (full time/part time/unemployed).
18. Employment category (paid/unpaid).
19. Hours employed per week (< 10/10–25/25–40/> 40).
20. Type of employment (unskilled manual/skilled manual/unskilled non-manual/skilled non-manual/professional/other/declined to answer).

BOX 2 List of candidate predictor variables from the baseline assessment data sets (*continued*)

21. Description of employment (if professional).
22. Description of employment (if other).
23. Occupation if not employed (retired/not looking for work/unable to work/looking for work/full-time student/other).
24. Description of unemployment (if other).
25. Education (CSE/O Level or GCSE/A level/degree/higher degree/other).
26. Description of level of education (if other).
27. Time on feet (most of the day/> 4 hours a day/< 4 hours a day/not much time, mostly sitting).
28. Time driving (most of the day/> 4 hours a day/< 4 hours a day/just to and from work/do not drive).
29. Current medications (since ankle injury/prior to injury/no/no answer).
30. Practice of physical activities **(11 questions)** (more than once per week/less than once per week/never).

41. Other physical activity (if other).
42. Height (cm).
43. Weight (kg).
44. Pain before injury (yes/no).
45. When had previous pain (during exercise or heavy activities, exercise and daily activities, constantly or other).
46. Description of when had previous pain (if other).
47. Frequency of previous pain (never/monthly/weekly/daily/always).
48. Previous instability (yes/no).
49. Severity of instability (mild/moderate/severe).
50. Frequency of instability (rarely/sometimes/frequently/always).
51. Previous injury (yes/no).
52. Three or more previous injuries (yes/no).
53. Previous injury < 1 year ago (yes/no).
54. Recurrent sprain – yes to all 3 questions above (yes/no).
55. ED attendance previously (yes/no).
56. How present injury occurred (during sport/at work/at home/outside in public place/other).
57. Description of how present injury occurred.
58. Maximum weight bearable (kg).

Baseline questionnaire

59. FAOS components **(42 questions)**.

101. Pain at rest VAS (0–100 points).
102. Pain bearing weight VAS (0–100 points).
103. FAOS baseline symptoms (subscale).
104. FAOS baseline pain (subscale).
105. FAOS baseline function ADL (subscale).
106. FAOS baseline function sport (subscale).
107. FAOS baseline QoL (subscale).
108. FLP components **(13 questions)**.

121. FLP work components **(10 questions)**.

131. FLP score.
132. FLP work score.
133. 1998 SF-12 components **(12 questions)**.

BOX 2 List of candidate predictor variables from the baseline assessment data sets (*continued*)

145. 1998 SF-12 physical score.
146. 1998 SF-12 mental score.
147. Baseline EQ-5D components **(5 questions)**.

152. Baseline EQ-5D score.
153. General level of health today (better/same/worse than the past 6 months).
154. VAS health today (0–100 points).

ADL, activities of daily living; A level, Advanced level; CSE, Certificate of Secondary Education; EQ-5D, EuroQol-5 Dimensions; FLP, Functional Limitation Profile; GCSE, General Certificate of Secondary Education; O level, Ordinary level; QoL, quality of life; SF-12, Short Form questionnaire-12 items; VAS, visual analogue scale.

Note
Imputed scores of validated scales with specific rules for handling missing data imputation (such as FAOS, SF-12 and EQ-5D) are also present in the CAST data set, but were not described here.

32 variables were preselected to form the group of candidate predictors considered to be plausibly predictive of either of the two outcomes. This initial selection was made internally by the research team, taking into account the results from the systematic review (see *Chapter 3*) and the conclusions from the consensus group meeting (see *Chapter 4*). The preselected candidate predictor variables and their details (type, name, categories or units, questionnaire in which the data were originally recorded and number of missing data) are listed in *Table 12*.

In addition to the baseline predictors, a few variables from the CAST 4-week follow-up questionnaire were selected to be investigated as potential predictors that could add some incremental value to the developed prognostic models. The list of these variables and their characteristics are listed in *Table 13*.

TABLE 12 Preselected candidate predictor variables from ED presentation and baseline assessment

Type	Variable name	Categories/units	Questionnaire	Missing values, *n* (%)
Binary	Sex	Male, female	Background information	0 (0)
	Previous pain	Yes, no	Background information	26 (4)
	Recurrent sprain	Yes, no	Background information	12 (2)
Categorical (or ordinal)	Employment status	No, part time, full time	Background information	0 (0)
	Education	CSE, GCSE, A level, degree, higher degree	Background information	20 (3)
	Anterior drawer test	Positive, painful, negative, no response	ED pro forma	396 (68)
	Talar tilt test	Positive, painful, negative, no response	ED pro forma	403 (69)
	Proximal fibular tender ligament test	Positive, painful, negative, no response	ED pro forma	378 (65)
	Able to bear weight	Full/partial/none	ED pro forma	322 (55)
	Treatment group	Tubular bandage, below-knee cast, Aircast brace, Bledsoe boot		0 (0)

TABLE 12 Preselected candidate predictor variables from ED presentation and baseline assessment (*continued*)

Type	Variable name	Categories/units	Questionnaire	Missing values, *n* (%)
	Leisure-time physical activity	None, < 1 time weekly, > 1 time weekly	Background information	7 (1)
	Walking ≥ 2 miles	None, < 1 time weekly, > 1 time weekly	Background information	24 (4)
	Previous instability	None, mild, moderate, severe	Background information	27 (5)
	Previous instability frequency	Never, rarely, sometimes, frequently, always	Background information	29 (5)
	Injury presentation	During sport, at work, at home, outside in public	Background information	34 (6)
	Ankle/foot swelling[a]	Never, rarely, sometimes, often, always	Baseline questionnaire	18 (3)
	Ankle/foot grinding/ clicking[a]	Never, rarely, sometimes, often, always	Baseline questionnaire	18 (3)
	Ankle/foot catching/ locking[a]	Never, rarely, sometimes, often, always	Baseline questionnaire	18 (3)
	Ankle ROM plantar flexion[a]	Never, rarely, sometimes, often, always	Baseline questionnaire	18 (3)
	Ankle ROM dorsiflexion[a]	Never, rarely, sometimes, often, always	Baseline questionnaire	18 (3)
	Pain at night (in bed)[a]	None, mild, moderate, severe, extreme	Baseline questionnaire	18 (3)
	Difficulty with squatting[a]	None, mild, moderate, severe, extreme	Baseline questionnaire	29 (5)
	Difficulty with running[a]	None, mild, moderate, severe, extreme	Baseline questionnaire	31 (5)
	Difficulty with jumping[a]	None, mild, moderate, severe, extreme	Baseline questionnaire	31 (5)
	Difficulty with twisting/ pivoting[a]	None, mild, moderate, severe, extreme	Baseline questionnaire	26 (4)
Continuous (or discrete)	Days from injury to assessment	0–7 days	ED pro forma/ background information	312 (55)
	Age	Years[b]	Background information	0 (0)
	BMI[c]	kg/m^2	Background information	19 (3)
	Maximum weight bearable	kg	Background information	5 (1)
	Pain when resting	VAS (0–100 points)	Baseline questionnaire	4 (1)
	Pain when bearing weight	VAS (0–100 points)	Baseline questionnaire	9 (2)
	SF-12 mental component	Score (0–100)	Baseline questionnaire	5 (1)

A level, Advanced level; CSE, Certificate of Secondary Education; GCSE, General Certificate of Secondary Education; ROM, range of motion; SF-12, Short Form questionnaire-12 items; VAS, visual analogue scale.
a Question from FAOS.
b An inclusion criteria for CAST was presenting at an ED to treat the ankle sprain no more than 7 days after the injury.
c Calculated from height and weight (both continuous variables), as collected in the CAST baseline questionnaire.

TABLE 13 Selected candidate predictor variables from the CAST 4-week follow-up questionnaire

Type	Variable name	Categories/units	Missing values, *n* (%)
Binary	Repeat injury to the same ankle	Yes, no	118 (20)
	Returned to ED because of repeated injury	Yes, no	120 (21)
Ordinal	Returned to usual sports/activities	No, partially, fully	121 (21)
	Ankle/foot swelling	Never, rarely, sometimes, often, always	102 (17)
	Ankle/foot grinding/clicking	Never, rarely, sometimes, often, always	102 (17)
	Ankle/foot catching/locking	Never, rarely, sometimes, often, always	103 (18)
	Able to perform ankle ROM plantar flexion	Never, rarely, sometimes, often, always	102 (17)
	Able to perform ankle ROM dorsiflexion	Never, rarely, sometimes, often, always	102 (17)
	Pain at night	None, mild, moderate, severe, extreme	101 (17)
	Difficulty with squatting	None, mild, moderate, severe, extreme	101 (17)
	Difficulty with running	None, mild, moderate, severe, extreme	135 (22)
	Difficulty with jumping	None, mild, moderate, severe, extreme	137 (23)
	Difficulty with twisting/pivoting	None, mild, moderate, severe, extreme	131 (22)
Continuous	Pain at weight bearing	0–100	196 (34)

ROM, range of motion.

Development data set preparation

The CAST data set contains individual participant information at baseline and at follow-up assessments (time points 2 and 3, as described in *Table 10*) for 584 participants recruited to take part in the study. To include the data collected at the time of ED presentation (time point 1, as described in *Table 10*) in the analysis, it was necessary to merge the CAST main data set with a separate data set that included information on 1487 people screened during the recruitment period of the trial. The information collected at this time point was anonymised; consequently, the data set has no information on the participants' identification number. Information from these two data sets was merged by matching the cases using the individuals' information on date of birth and sex; any duplicates in each data set were disregarded to avoid mismatching.

This process added information on five of the candidate predictors collected during the first contact with the participants to 289 cases in the CAST main data set (see *Box 1* for details on the predictors collected at the time of ED presentation). As the results of the trial have been published and all documentation archived, there was no need for further data cleaning and the only data manipulation performed with the resulting data set was generating new variables from existing variables or recategorisation of existing variables (see *Available candidate predictors and initial selection of variables for modelling* and *Exploratory analysis and data transformation*) and missing data imputation (see *Handling missing data*).

Exploratory analysis and data transformation

Baseline and 4-week follow-up characteristics of the participants in CAST were summarised using means, standard deviations (SDs) and ranges for continuous variables, or counts and percentages for categorical variables. After merging the data sets, three variables from the ED data set had > 60% missing information (anterior drawer test, talar tilt test and proximal fibular tender ligament test) and were excluded from the list of candidate predictors (see *Table 12* for detailed information on the number of missing data for each candidate predictor). It was also discussed and agreed during the consensus group meeting (see *Chapter 4*)

that it would be reasonable to exclude these variables from the pool of candidate predictors because of the variability in technique between assessors when performing the tests.

Each binary or categorical predictor was tabulated against the outcomes to check for empty or low cell counts. When this was the case, categorical variables were recategorised by collapsing some of their categories, providing it made clinical sense to do so. The manipulated variables with details on the changes performed are presented in *Table 14*. When collapsing categories would not solve the problem (or make clinical sense), the predictor variable was omitted from any further analysis. This was the case for the following candidate predictors: 'ankle/foot swelling' (from baseline assessment) and 'returned to ED because of repeat injury' (from 4-week follow-up questionnaire).

The distribution of continuous predictors was also assessed: first, by considering the predictors' empirical distributions by producing histograms, and then by assessing these for normality by means of normal probability plots. The presence of any outliers was assessed based on visual examination of box plots. Extreme values were inspected to confirm whether or not they were clinically plausible. No individual participant information was deleted from the data set and data transformation (normalisation) was performed as appropriate.

The correlations between candidate predictors were also examined using Spearman's rank-correlation coefficient to identify any highly correlated predictors ($r \geq 0.8$). It causes unnecessary complication to include highly correlated predictors together in multivariable models. Highly correlated predictors explain the same variation in outcome, and this was found for two groups of variables: (1) 'difficulty with running', 'difficulty with jumping' and 'difficulty with twisting/pivoting' (from both baseline and 4-week follow-up) and (2) 'previous instability' and 'previous instability frequency' (from baseline).

To deal with the first group of correlated variables, a new binary variable (yes or no) was created to indicate whether or not a participant presented a positive answer to any of the original variables; these individuals were characterised as presenting difficulty with running, jumping or twisting. This new composite variable was then used instead of the three highly correlated variables in the remaining analyses. The decision about which predictor should be taken to the modelling stage, between previous instability and previous instability frequency, took into account the individual predictive ability of each variable. The predictor with lower face validity for outcomes 1 and 2 ('previous instability' in both cases) was then omitted from subsequent analyses.

Initial individual associations between each candidate predictor and poor recovery at 9 months after ankle sprain were performed by fitting unadjusted logistic regression models for outcomes 1 and 2.

Handling missing data

Some missing data in the development data set occurred as a result of missed appointments and losses to follow-up during the conduct of CAST, but also because of the lack of a unique patient identification in the trial's screening (ED presentation) data set, which did not allow all the information collected at this point to be merged with the main CAST data set (see *Development data set preparation* for further details). The percentage of missing data in the final merged data set is presented for each candidate prognostic variable in *Tables 12* and *13*. To conform to current guidelines, multiple imputation for all participants with at least one missing value was performed.[53] Since there were several predictor variables of different types (i.e. binary, categorical and continuous) with missing data, multiple imputation by chained equations (MICE) was carried out using the *mi impute chained* function in Stata® version 14.2 (StataCorp, College Station, TX, USA) with the options *logit* (for imputation of binary variables), *mlogit* (for imputation of categorical variables) and *truncreg* (for imputation of continuous variables, setting the lower and upper limits for imputed values as 0 and 100, respectively).

TABLE 14 Format and categories/units of candidate predictor variables in the original CAST data set and after data manipulation

	In the original data set		After exploratory analysis/data manipulation	
Variable name	Type	Categories/units	Type	Categories/units
Employment status	Categorical	None, part time, full time, student, retired	Categorical	None,[a] part time, full time
Injury presentation		During sport, at work, at home, in public, other[b]		During sport, at work, at home, in public
Leisure time physical activities (several types of activities)[c]		None, < 1 time weekly, > 1 time weekly		None, < 1 time weekly, > 1 time weekly
Ankle/foot catching/ locking		Never, rarely, sometimes, often, always		Never, rarely/sometimes, often/always
Ankle/foot grinding/ clicking		Never, rarely, sometimes, often, always		Never, rarely/sometimes, often/always
Previous instability frequency		Never, rarely, sometimes, often, always		Never, rarely/sometimes, often/always
Able to perform ankle ROM plantarflexion		Always, often, sometimes, rarely, never		Often/always, rarely/ sometimes, never
Able to perform ankle ROM dorsiflexion		Always, often, sometimes, rarely, never		Often/always, rarely/ sometimes, never
Pain at night (in bed)		None, mild, moderate, severe, extreme		None/mild/moderate, severe/extreme
Difficulty with squatting		None, mild, moderate, severe, extreme		None/mild/moderate, severe/extreme
Difficulty with running		None, mild, moderate, severe, extreme		None/mild/moderate, severe/extreme
Difficulty with jumping		None, mild, moderate, severe, extreme		
Difficulty with twisting/ pivoting		None, mild, moderate, severe, extreme		
Anterior drawer test		Positive, painful, negative	Binary	Positive/painful, negative
Talar tilt test		Positive, painful, negative		Positive/painful, negative
Proximal fibular tender ligament test		Positive, painful, negative		Positive/painful, negative
Weight-bearing ability (at ED presentation)		Full, partial, none		Yes,[d] no
Days from injury to assessment[e]		Days		• 1–2 days • 3–7 days
Maximum weight bearable (baseline assessment)		kg		• Unable to perform test • Able to bear some weight

ROM, range of motion.
a Combination of unemployed, student and retired.
b The answers under the option 'other' were reviewed and regrouped with the remaining options accordingly.
c From a list with 11 different activities (plus an option for any other activity that might not be covered) that were combined into a single variable indicating the highest frequency reported for any physical activity (except 'walking ≥ 2 miles' which was explored separately).
d Combination of full and partial ability to bear weight.
e Not allowed > 7 days in the CAST data set.

In MICE, all missing values are filled in by simple random sampling with replacement from the observed values to allow the regression models to be fitted on all values. Then, the variable with the lowest number of missing observations, for example x_1, is regressed on all other variables. Missing values are then replaced by drawing from the estimated corresponding posterior predictive distribution of x_1. Then, the next variable with the lowest number of missing observations is regressed on all other variables including (and using the imputed values of) x_1. This process is repeated until all variables with missing values are imputed, forming one cycle. Cycles are repeated to stabilise the results and the whole procedure is repeated m times to give m imputed data sets. An important characteristic of MICE is the capacity of handling different variable types (continuous, binary, unordered and ordered categorical) because each variable is imputed using its own imputation model using different types of regression analysis.

Multiple imputation was performed under the assumption that all missing data were missing at random (MAR). In other words, the probability of data being missing does not depend on the unobserved data, conditional on the observed data. Therefore, imputation models included all available observed characteristics for the predictors of interest (both at baseline and at the 4-week follow-up), predictors of predictors (e.g. weight and height for BMI) and the outcomes, as recommended by White *et al.*[53] The models were independently estimated for outcomes 1 and 2, and imputations were therefore performed in separate procedures, producing two different sets of 50 complete data sets. This number of imputed data sets was chosen based on the number of missing data for the variable with the highest rate of missing observations (312/584 for 'days from injury to assessment'). No data transformation was performed on continuous predictor variables before imputing missing observations.

Despite using the augmented-regression approach,[54] some predictors were also excluded during this process because of the issue of 'perfect prediction' when imputing categorical variables.[55] Perfect prediction occurs whenever there is a level of a categorical explanatory variable for which the observed values of the outcome are all 1 (or all 0). Perfect prediction then leads to infinite coefficients with infinite standard errors and causes instability during estimation, which prevents the imputation model from achieving convergence. We resolved this issue by dropping the predictors causing the perfect prediction from the multiple imputation model: two from baseline [(1) 'difficulty with running, jumping or twisting' and (2) 'previous instability frequency'] and one from 4-week follow-up ['able to perform ankle range of motion plantarflexion']. A complete list of predictors that were excluded before the modelling process, with reasons for exclusion, is provided in *Table 15*.

Sample size considerations

Sample size requirements for logistic regression are based on the concept of events per variable (EPV). It is widely recommended that, to develop a prediction model, the data set should contain a minimum of 5–10 EPV.[56–61] Based on a number of at least five EPV, the outcome rates (see *Table 11*) observed in the CAST data set allowed the inclusion of 23 (116/5) and 28 (140/5) candidate predictor variables in the models for outcomes 1 and 2, respectively. After the exclusion of nine preselected candidate predictors for the reasons described in *Table 15*, 23 variables from baseline remained as candidate predictors. However, some of these predictors were categorical variables with more than two levels, which affects the EPV as these predictors require the generation of indicator variables for each category (e.g. employment status coded as 'no', 'part time' or 'full time' will require three parameters to be estimated). Therefore, we ended with 35 candidate parameters, which means that the EPV ratio was approximately 3 and 4 for outcomes 1 and 2, respectively. It is also important to note that some of the candidate predictors were continuous variables, which could require non-linear modelling and therefore increase even more the number of regression coefficients to be estimated and affect the EPV (e.g. if using fractional polynomials and the best transformation for age was found to be age + age^2, then age would relate to two predictors instead of one). However, this was not the case.

To the best of our knowledge, this was the first project aiming to develop prediction models to assess the risk of poor recovery after an acute ankle sprain. Therefore, we have opted for relaxing the EPV rule in favour of including more potentially important predictors in the analyses. However, we have adopted

TABLE 15 Reason for exclusion of predictors before the modelling process

Predictor	Reason for exclusion
Baseline	
Anterior drawer test	≥ 60% missing values, consensus agreement
Talar tilt test	≥ 60% missing values, consensus agreement
Proximal fibular tender ligament test	≥ 60% missing values, consensus agreement
Ankle/foot swelling	One or more cells with too few cases when cross-tabulated with the outcomes, regardless of recategorisation
Difficulty with running	Highly correlated with 'difficulty with jumping' and 'difficulty with twisting/pivoting'. Composite variable used instead
Difficulty with jumping	Highly correlated with 'difficulty with running' and 'difficulty with twisting/pivoting'. Composite variable used instead
Difficulty with twisting/pivoting	Highly correlated with 'difficulty with running' and 'difficulty with jumping'. Composite variable used instead
Previous instability	Highly correlated with 'previous instability frequency'
Previous instability frequency	'Perfect prediction' during missing data multiple imputation
Difficulty with running/jumping/twisting	'Perfect prediction' during missing data multiple imputation
4-week follow-up	
Returned to ED because of repeated injury	One or more cells with too few cases when cross-tabulated with the outcomes
Difficulty with running	Highly correlated with 'difficulty with jumping' and 'difficulty with twisting/pivoting'. Composite variable used instead
Difficulty with jumping	Highly correlated with 'difficulty with running' and 'difficulty with twisting/pivoting'. Composite variable used instead
Difficulty with twisting/pivoting	Highly correlated with 'difficulty with running' and 'difficulty with jumping'. Composite variable used instead
Ankle ROM dorsiflexion	'Perfect prediction' during missing data multiple imputation

ROM, range of motion.
Note
Perfect prediction occurs whenever there is a level of a categorical explanatory variable for which the observed values of the outcome are all 1 (or all 0); it is often resolved by discarding the observations corresponding to offending covariate patterns or the independent variables perfectly predicting outcomes during estimation.

several strategies to minimise bias and overfitting, including the estimation of heuristic shrinkage factors to account for possible extreme predictions resulting from overestimated associations (see *Data modelling*, *Model update*, *Assessment of model performance* and *Shrinkage*).

Data modelling

Since both outcomes were binary (poor outcome after ankle sprain – yes/no), the prognostic models were developed using a logistic regression modelling framework with the logit probability of poor outcome as the response variable. The 23 remaining candidate predictors were included together in full logistic regression models as independent variables, and further selection of predictors was based on the statistical significance of their adjusted relationship with the outcomes. At this point, continuous variables were kept as continuous to avoid loss of prognostic information.[62] Therefore, the shape of the relationship between continuous predictors and the outcome should be studied and modelling performed with non-linear functions, such as fractional polynomials, when appropriate.[63]

Non-linear relationships were investigated using fractional polynomials and the 'best transformation' for each continuous predictor was used when fitting the models. As more than one continuous variable was included in the full models, the multivariable fractional polynomial (MFP) algorithm was used.[64,65] The MFP algorithm selects predictors and their transformations that best predict the outcome variable using a backward selection process. A nominal alpha of 0.15 was used to warrant exclusion from the model to reduce the risk of overfitting. Another advantage of the MFP algorithm is that selection of predictors and transformations is done simultaneously, preserving the nominal type 1 statistical error probability.

As the analyses were performed in sets of 50 multiply imputed data sets; the MFP algorithm was applied using the Stata command *mfpmi* together with *logit*. The *mfpmi* allows binary, ordinal and non-ordinal categorical variables to be included alongside continuous variables in the same model, and simultaneously select the appropriate fractional polynomial transformation of continuous predictors combining the estimates of multiply imputed data sets. The multivariable models were fitted in each of the 50 complete data sets and the estimated regression parameters (coefficients and variances) were combined using Rubin's rule.[66,67]

Ideally, prognostic models should be flexible, easy to understand and parsimonious, so that they are simple and quick to apply in clinical practice. Therefore, after identifying the best transformation terms for continuous variables in the full multivariable models with all candidate predictors, the statistically significant predictors (and the corresponding transformations of continuous variables, when applicable) were selected using the AIC as the decision rule and kept in the final model.[68] Therefore, a p-value of < 0.157 (equivalent to AIC) was conservatively taken to warrant inclusion of predictors in the final model and to reduce the risk of overfitting.

Model update

After developing the prognostic models for outcomes 1 and 2 including only predictors collected at baseline (baseline variables), the additional incremental value of candidate predictors collected at the 4-week follow-up point were investigated. First, all additional candidate predictors were included together in the final baseline models and only those predictors achieving $p < 0.157$ (AIC) were considered for inclusion in the updated models (i.e. prognostic models including baseline + 4-week predictors). Finally, these updated models were compared with the original baseline models by DCA plots[69,70] to investigate whether or not the inclusion of additional predictors was reflected in an increased net benefit. The DCA was performed in Stata, using the command *dca*.

Assessment of model performance

After developing a prognostic model, it is important to evaluate its performance. *Table 16* provides an overview of the main ways in which model performance can be assessed from Thangaratinam *et al.*[71]

The performance of the prognostic models was characterised by evaluating calibration and discrimination.

Calibration

Calibration is the agreement between observed and predicted probabilities of poor outcome. The calibration of the developed prognostic models was assessed graphically using calibration plots, with observed risks plotted on the y-axis against predicted risks on the x-axis.[72,73] The calibration plot is created by regressing the occurrence of the outcome on the predicted probability of the outcome using locally weighted scatterplot smoothing (LOWESS). This plot shows the direction and magnitude of model miscalibration across the probability range. The calibration plot was also supplemented with estimates of the calibration slope and intercept. Models with perfect calibration will have a calibration slope of 1 and intercept 0 (i.e. prediction lying on or around the 45° line).

Discrimination

Discrimination is the ability of the prognostic model to separate individuals with the outcome from those without (i.e. those with the outcome should have higher predicted probabilities than those without). The overall discriminatory ability was summarised by the c-statistic (or area under receiver operating

TABLE 16 Main methods of assessing prognostic model performance

Terms	Definitions
Calibration	Calibration indicates the ability of the model to correctly estimate the absolute risks and was examined using calibration plots
Reproducibility (internal validation)	The process of determining internal validity. Internal validation assesses validity for the setting from which the development data originated
Generalisability/transportability (external validation)	The process of determining external validity of the prediction model to populations that are plausibly related
Discrimination	Discrimination describes the ability of the model to correctly distinguish those who will have an adverse outcome from those who will not
Calibration plot	In a calibration plot, the predictive risk is plotted against the observed incidence of the outcome. Ideally the predicted risk equals the observed incidence throughout the entire risk spectrum and the calibration plot follows the 45° line

characteristic curve) with 95% confidence interval (CI). The *c*-statistic was classified as follows: 0.5–0.6, fail; 0.6–0.7, poor; 0.7–0.8, fair; 0.8–0.9, good; 0.9–1.0, excellent.

Owing to complexities in the model building (e.g. a combination of variable selection, fractional polynomials and multiple imputation), we did not carry out an internal validation of the model (e.g. using bootstrapping), as not all these approaches could be replayed in the internal validation. We therefore carried out an ad hoc hybrid of apparent performance and internal validation, whereby model performance was evaluated both on the original CAST data and also separately in each imputed data set. We calculated the model discrimination in the original CAST data, and also combining the results obtained from multiply imputed data sets using Rubin's rules. Calibration plots were created following recommendations of overlaying calibration curves from each imputed data set.[74]

Shrinkage

Newly developed prognostic models are often optimistic as a result of overfitting, which leads to worse prediction in independent data. Reasons for overfitting include small EPV, the selection of predictors based on *p*-values and modelling non–linear relationships between predictors and the outcome. To estimate the amount of overfitting likely to be present in the developed prognostic models, heuristic shrinkage factors were calculated independently for each model as:

$$(\text{model } \chi^2 - \text{df})/\text{model } \chi^2, \quad (1)$$

where model χ^2 is the model likelihood ratio, or –2log-likelihood of a model with only an intercept and the fitted model, and df is the number of degrees of freedom in the fitted model. The number of degrees of freedom in the fitted model is defined by the number of degrees of freedom considered for all explored candidate predictors, plus all corresponding transformations, when applicable.

A shrinkage factor of 1 implies no shrinkage. The regression coefficients from the prognostic models were multiplied by the shrinkage factor to adjust the models for optimism. The shrinkage of the intercept was estimated by fitting a logistic regression model for each studied outcome, including the linear predictor (log-odds) calculated using the shrunk coefficients as the only independent variable, and constraining its coefficient to one (offset variable).[75]

Results

Baseline characteristics

The baseline characteristics of the participants in the CAST data set are summarised in *Table 17*. Participants were aged 29.88 years, on average, with the age range varying from 16 to 72 years. Participants had a mean BMI of 26.34 kg/m^2 and lower pain sores when resting (mean 37.75/100 points) than when bearing weight on the injured ankle (mean 75.42/100 points); < 25% of participants reported not being able to bear any weight on their ankles at the time of baseline assessment. Most participants reported not feeling pain in the ankle before the injury (86.56%) and not seeking treatment for a recurrent sprain (90.38%). Most participants were in full-time employment (61.64%), had an education level higher than General Certificate of Secondary Education (GCSE) (84.98%) and fewer than one-quarter engaged in any leisure-time physical activity more than once a week (24.09%). Among the CAST participants, injuries occurred mostly during the practice of sports (36.91%).

All continuous variables presented at least a minimal departure from a normal distribution, as evidenced in *Figures 4–10*. Some outliers were observed for participants' age, weight, BMI and pain score when bearing weight. However, all extreme values were clinically plausible, so no observations were dismissed.

TABLE 17 Summary of baseline characteristics (candidate predictors) of the CAST sample

Variable	Mean (SD)	Minimum, maximum
Age (years)	29.88 (10.77)	16, 72
Height (m)	1.73 (0.98)	1.47, 2.01
Weight (kg)	78.56 (15.44)	39.92, 133.36
BMI (kg/m^2)	26.34 (5.19)	16.07, 53.77
Pain when resting (score), points	37.75 (23.49)	0, 100
Pain when bearing weight (score), points	75.42 (19.61)	0, 100
SF-12 Mental Component (score), points	51.08 (11.26)	20.55, 68.77
	Frequency	***%***
Sex		
Male	337	57.71
Female	247	42.29
Days from injury to assessment		
0–2	118	44.87
≥ 3	145	55.13
Able to bear weight at ED presentation		
No	72	27.48
Yes	190	72.52
Able to bear weight at baseline assessment		
No	446	77.03
Yes	133	22.97
Pain on the ankle before injury		
No	483	86.56
Yes	75	13.44

continued

TABLE 17 Summary of baseline characteristics (candidate predictors) of the CAST sample (*continued*)

Variable	Mean (SD)	Minimum, maximum
Recurrent sprain		
No	517	90.38
Yes	55	9.62
Pain in bed at night		
No	378	66.78
Yes	188	33.22
Difficulty with squatting		
None/mild/moderate	88	15.86
Severe/extreme	467	84.14
Current employment		
None	132	22.6
Part time	92	15.75
Full time	360	61.65
Treatment received for ankle sprain		
Tubular bandage	144	24.66
Below-knee cast	142	24.32
Aircast brace	149	25.51
Bledsoe boot	149	25.51
Education level		
CSE level or lower	84	15.02
O level/GCSE/A level	383	68.52
Degree/higher degree	92	16.46
Leisure-time physical activity		
None	28	4.85
< 1 time weekly	410	71.06
> 1 time weekly	139	24.09
Walking ≥ 2 miles per day		
None	164	29.29
< 1 time weekly	105	18.75
> 1 time weekly	291	51.96
Injury mechanism		
At home	99	18.00
Practising sports	203	36.91
At work	79	14.36
Outside, in public	169	30.73
Ankle grinding/clicking		
Never	257	45.41
Rarely/sometimes	220	38.87
Often/always	89	15.72

TABLE 17 Summary of baseline characteristics (candidate predictors) of the CAST sample (*continued*)

Variable	Mean (SD)	Minimum, maximum
Ankle catching/locking		
Never	286	50.53
Rarely/sometimes	209	36.93
Often/always	71	12.54
Ankle ROM plantar flexion		
Always/often	101	17.84
Sometimes/rarely	247	43.64
Never	218	38.52
Ankle ROM dorsiflexion		
Always/often	81	14.31
Sometimes/rarely	227	40.11
Never	258	45.58

A level, Advanced level; CSE, Certificate of Secondary Education; O level, Ordinary level; ROM, range of motion; SF-12, Short Form questionnaire-12 items.

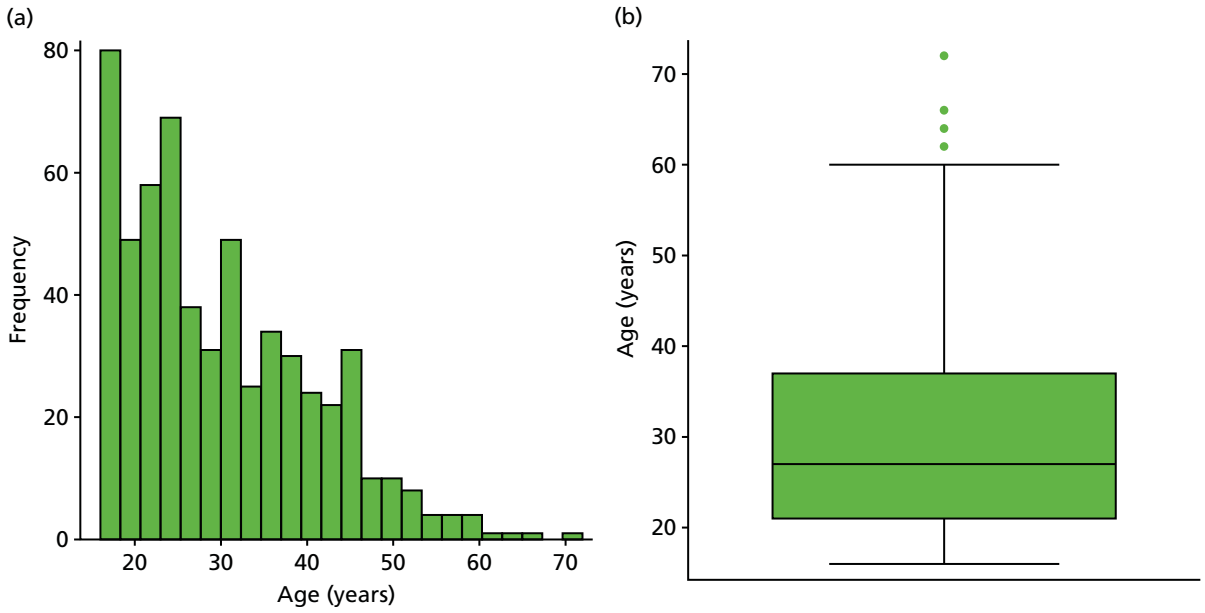


FIGURE 4 Participants' age. (a) Histogram; and (b) box plot.

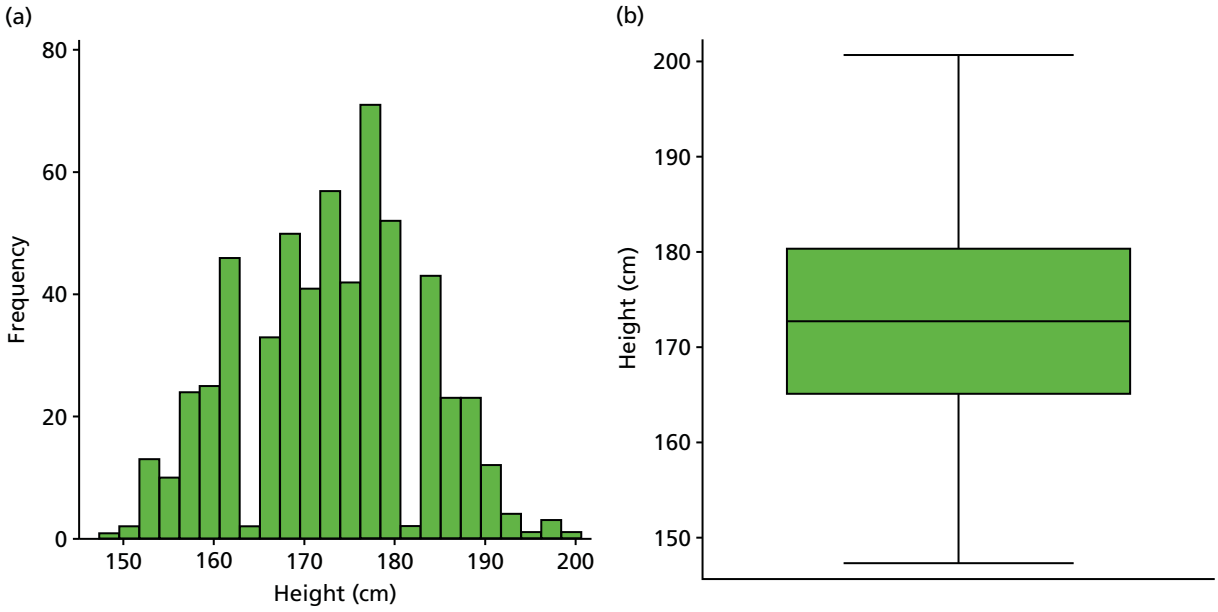


FIGURE 5 Participants' height. (a) Histogram; and (b) box plot.

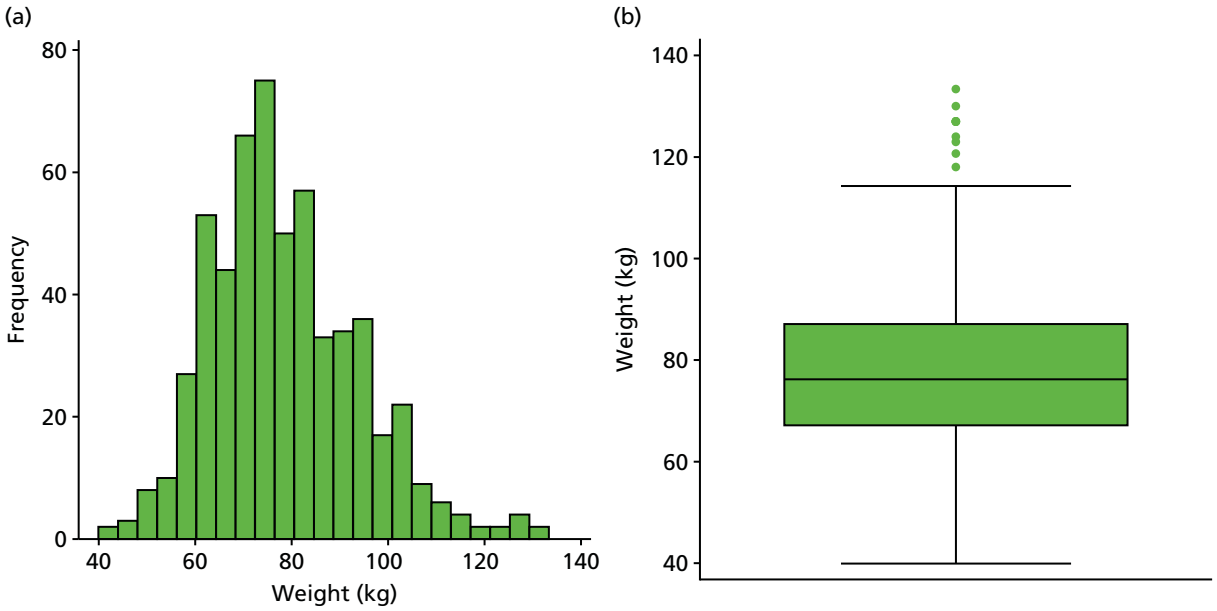


FIGURE 6 Participants' weight. (a) Histogram; and (b) box plot.

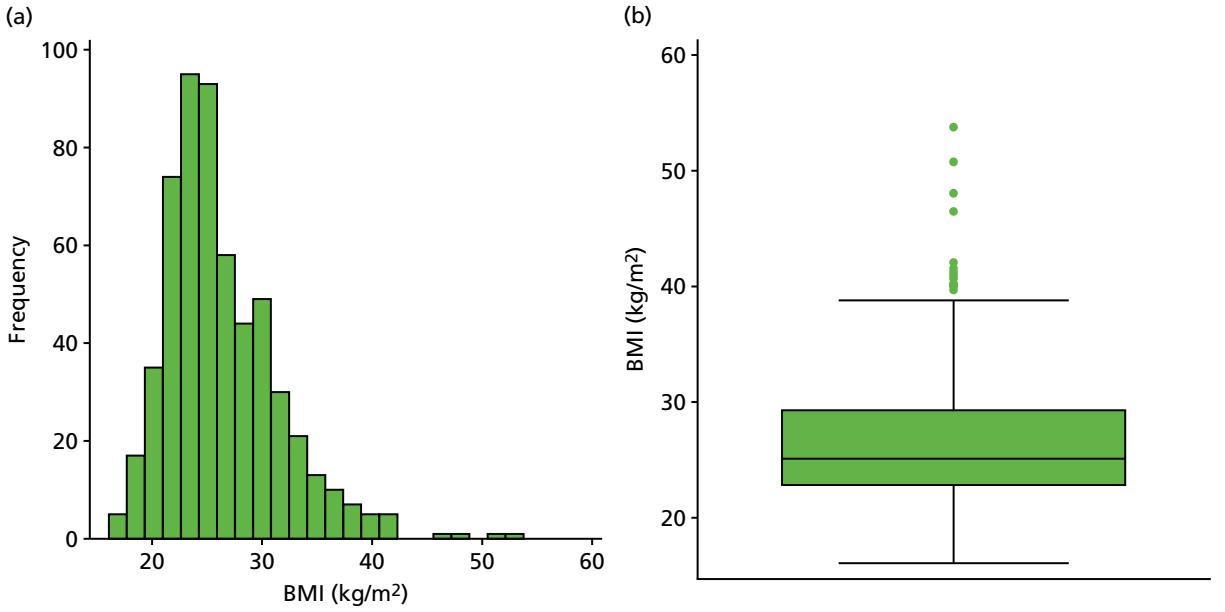


FIGURE 7 Participants' BMI. (a) Histogram; and (b) box plot.

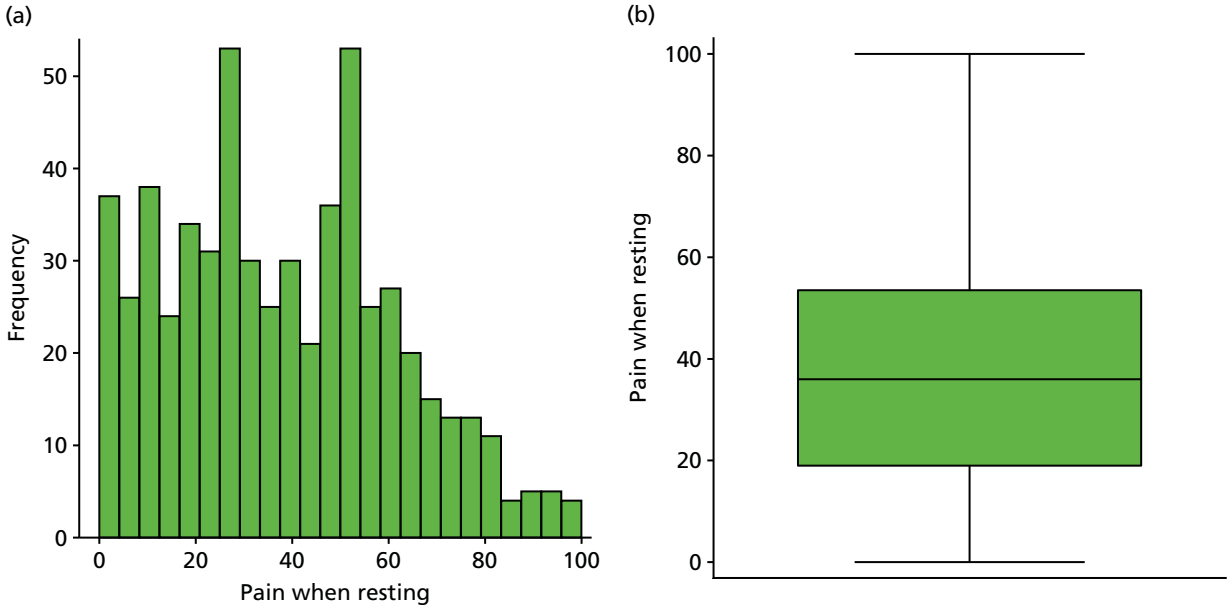


FIGURE 8 Pain score when resting at ED presentation. (a) Histogram; and (b) box plot.

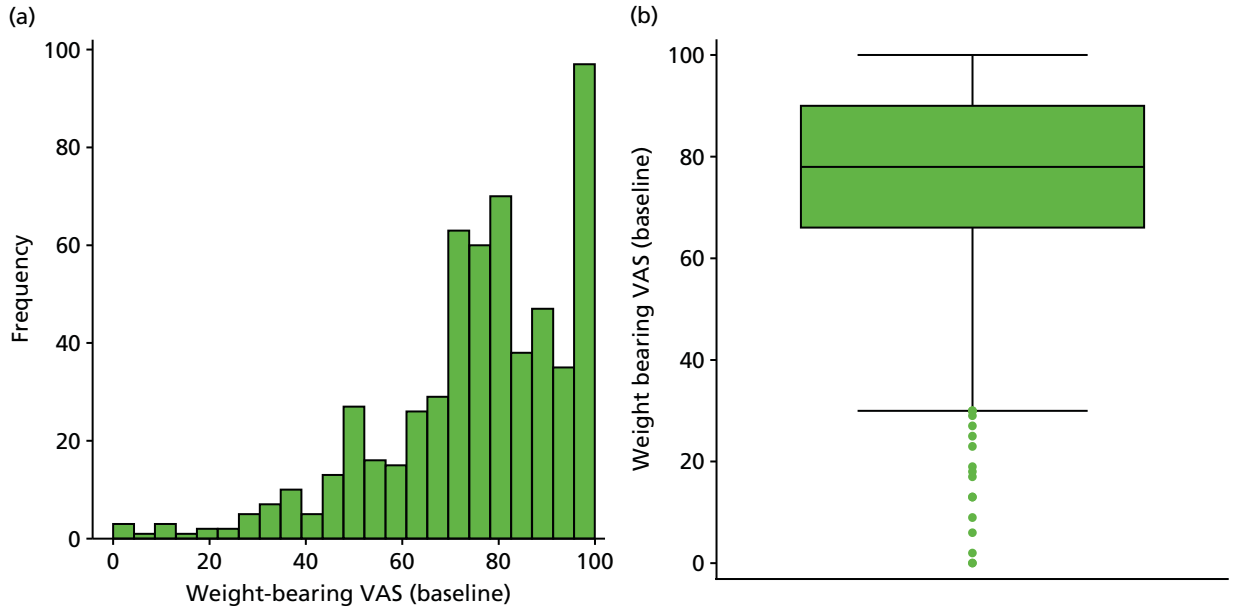


FIGURE 9 Pain score when bearing weight at baseline assessment. (a) Histogram; and (b) box plot. VAS, visual analogue scale.

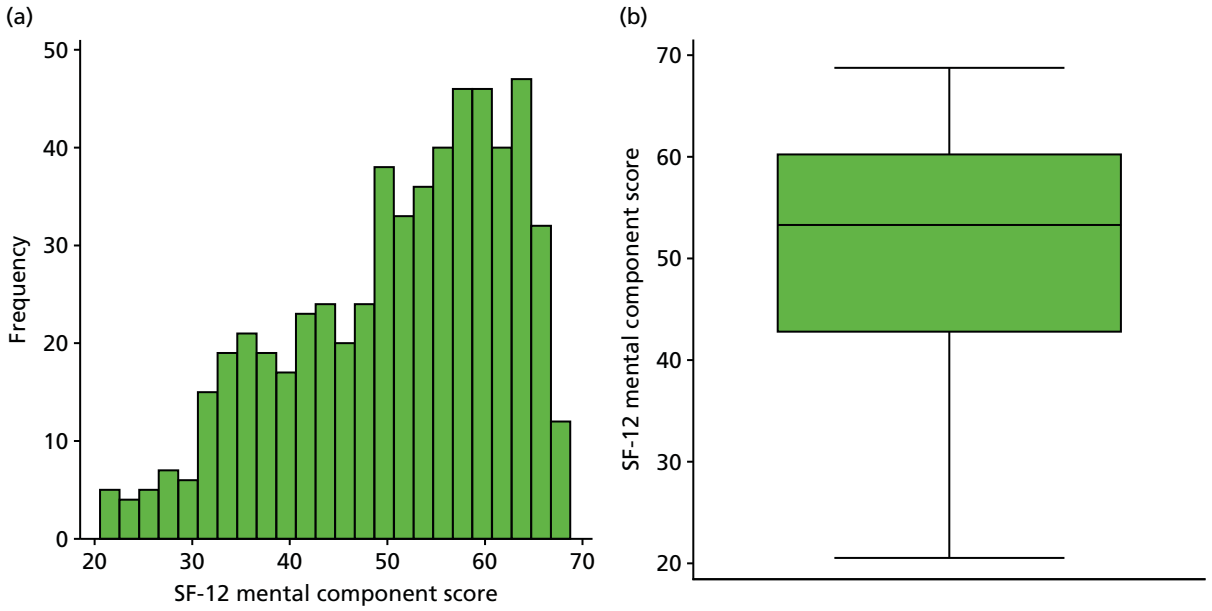


FIGURE 10 Short Form questionnaire-12 items (SF-12) mental component score. (a) Histogram; and (b) box plot.

Spearman's rank-correlation coefficients of the baseline predictors are presented in *Table 18*. Highly correlated candidate predictors included 'difficulty with running' and 'difficulty with jumping' ($r = 1.000$); 'difficulty with running' and 'difficulty with twisting/pivoting' ($r = 0.859$); 'difficulty with jumping' and 'difficulty with twisting/pivoting' ($r = 0.859$); and 'previous instability' and 'previous instability frequency' ($r = 0.997$). As these variables should not be included together in the regression models, the first set of highly correlated variables was combined into a single composite variable to identify those participants with difficulties in running, jumping or twisting/pivoting. For the second pair of highly correlated variables, previous instability frequency only was included in the subsequent analysis.

Multivariable models

The summary of the full multivariable model estimates (predictor coefficients, 95% CIs and *p*-values) is presented in *Table 19*. For outcome 1, 7 of the 23 candidate predictors were selected for inclusion in the final model, based on the AIC ($p < 0.157$): (1) age, (2) BMI, (3) pain when resting, (4) pain when bearing weight, (5) number of days from injury to assessment, (6) ability to bear weight and (7) whether or not the injury was a recurrent sprain. For the outcome 2, almost the same set of candidate predictors were selected for inclusion in the final model, except for age and BMI. For outcome 2 educational level was found to be a statistically important candidate predictor. However, education was identified as a low-priority variable by the consensus committee. There were particular difficulties with this variable, as the criteria used in the CAST study to identify different education achievements have been superseded, and in the interim, a number of new additional categories of study have become more popular (for example University of the Third Age). Given the marginal statistical significance, inability to replicate the categories in an external validation, low priority given in the consensus, and the reluctance of clinicians to probe this information, we did not include this variable in the final model for outcome 2.

The best fit for all continuous predictors was found to be linear transformations (mean subtractions), which were incorporated into the model by updating the intercepts accordingly. A summary of the estimates from the final multivariable models (predictor coefficients, 95% CIs and *p*-values) is presented in *Table 20*. For outcome 1, BMI was not statistically significant according to AIC in the final model. Nevertheless, it was decided not to exclude this variable from the model, given its clinical importance, and to reduce the risk of overfitting. Both models were fairly simple, composed of just a few predictors that are routinely collected in the clinical setting.

Only pain when bearing weight at 4 weeks after the injury was included in the updated models (baseline + week 4 predictors) for both outcomes 1 and 2 (*Table 21*). By inspecting the DCA plots shown in *Figures 11* and *12*, it is possible to see a clear net benefit gain over the entire range of thresholds when using any of the developed prognostic models in comparison to considering all patients (or no patient) at risk of having poor outcome after an acute ankle sprain. Furthermore, the inclusion of the week 4 predictor (pain when bearing weight) consistently improved the performance of the models for both outcomes 1 and 2.

Model performance

Model performance was assessed in terms of calibration and discrimination. The overall discriminatory ability (apparent performance) was 0.82 (95% CI 0.75 to 0.89) for the model developed to predict outcome 1 and 0.73 (95% CI 0.66 to 0.81) for the model developed to predict outcome 2, as measured by the *c*-statistic estimated after regressing the predictors selected for the final model against the outcomes using the original CAST data set (complete-case analysis, $n = 194$ and $n = 200$ for outcomes 1 and 2, respectively). The combined results from the analysis of the 50 imputed data sets provided a less optimistic measure of the discriminatory ability for the two models. For the model developed to predict outcome 1, the combined *c*-statistic was 0.74 (95% CI 0.70 to 0.79). For the model developed to predict outcome 2, the combined *c*-statistic was 0.70 (95% CI 0.65 to 0.74). The addition of one variable with information on pain when bearing weight on the ankle at 4 weeks after the injury improved the discriminatory ability and apparent calibration of both models. For the updated model to predict outcome 1, the *c*-statistic was 0.77 (95% CI 0.73 to 0.82). For the updated model to predict outcome 2, the *c*-statistic was 0.75 (95% CI 0.71 to 0.80).

TABLE 18 Spearman's rank-correlation coefficients between pre-selected candidate predictors from baseline

	Days from injury to assessment	Maximum bearable weight	Sex	Pain before injury	Recurrent sprain	Current employment
Days from injury to assessment	–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Maximum bearable weight	0.050	–	*ref.*	*ref.*	*ref.*	*ref.*
Sex	0.014	0.019	–	*ref.*	*ref.*	*ref.*
Pain before injury	–0.039	–0.026	0.112	–	*ref.*	*ref.*
Recurrent sprain	–0.008	–0.008	0.042	0.389	–	*ref.*
Current employment	0.046	–0.053	–0.438	–0.065	–0.016	–
Education	–0.085	–0.049	–0.075	–0.070	–0.055	0.204
Treatment arm	–0.047	0.199	–0.037	–0.016	–0.024	–0.033
LTPA	0.000	0.003	0.166	0.067	–0.068	–0.138
Walking	0.074	0.062	0.214	0.079	0.028	–0.123
Previous instability	–0.056	–0.011	0.080	0.470	0.331	–0.174
Previous instability frequency	–0.043	–0.014	0.079	0.464	0.322	–0.179
Injury mechanism	0.110	0.005	0.409	0.026	0.008	–0.288
Ankle grinding	0.079	0.033	0.049	0.267	0.136	–0.033
Ankle catching or locking	0.104	0.048	–0.005	0.192	0.013	–0.060
Plantar ROM flexion	–0.085	0.182	–0.021	–0.020	–0.043	–0.058
Plantar ROM dorsiflexion	–0.186	0.147	–0.064	0.048	0.009	0.010
Pain at night	–0.119	0.062	0.201	0.086	–0.041	–0.193
Difficulty with squatting	–0.011	0.145	0.049	–0.199	–0.110	–0.054
Difficulty with running	0.005	0.086	0.020	–0.128	–0.102	0.027
Difficulty with jumping	0.005	0.086	0.020	–0.128	–0.102	0.027
Difficulty with twisting	–0.050	0.178	0.066	–0.088	–0.070	0.014
Age	0.101	–0.137	0.127	–0.022	–0.054	0.048
BMI	0.049	–0.123	0.268	0.113	0.010	–0.135
Able to bear weight	0.204	0.149	0.082	0.029	–0.001	–0.041
Pain when resting	–0.066	0.072	0.246	0.075	0.028	–0.234
Pain when bearing weight	–0.144	0.161	0.209	0.028	0.066	–0.196
SF-12 mental component	–0.002	0.115	–0.074	–0.153	–0.108	0.106

LTPA, Leisure Time Physical Activity; ROM, range of motion; SF-12, Short Form questionnaire-12 items.

Note

Variables excluded because of the amount of missing data are not included in this table.

Education	Treatment arm	LTPA	Walking	Previous instability	Previous instability frequency	Injury mechanism	Ankle grinding
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
–0.046	–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
0.055	–0.142	–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
–0.086	–0.045	0.053	–	*ref.*	*ref.*	*ref.*	*ref.*
–0.051	–0.115	0.015	–0.021	–	*ref.*	*ref.*	*ref.*
–0.045	–0.114	0.019	–0.014	0.997	–	*ref.*	*ref.*
–0.189	–0.054	0.063	0.113	0.066	0.065	–	*ref.*
–0.150	–0.037	–0.108	–0.053	0.211	0.217	–0.020	–
–0.164	0.050	0.063	–0.050	0.122	0.127	0.104	0.486
–0.014	–0.002	0.015	–0.016	0.010	0.009	0.009	–0.032
0.033	–0.021	0.032	0.063	0.006	0.006	–0.075	–0.110
–0.078	–0.009	0.128	–0.024	0.035	0.040	0.176	0.224
–0.008	0.040	–0.106	–0.042	–0.041	–0.036	0.021	0.078
0.080	–0.007	0.072	–0.091	–0.108	–0.093	0.037	0.043
0.080	–0.007	0.072	–0.091	–0.108	–0.093	0.037	0.043
0.032	0.081	–0.002	–0.091	–0.067	–0.054	0.051	0.103
–0.004	0.004	0.124	–0.025	0.019	0.029	0.174	–0.088
–0.064	–0.063	0.174	0.018	0.048	0.052	0.196	0.190
–0.042	0.007	0.001	0.205	–0.018	–0.013	–0.049	–0.031
–0.157	–0.003	0.103	–0.005	0.106	0.102	0.198	0.305
–0.133	0.078	0.097	–0.031	0.086	0.084	0.159	0.091
0.131	0.016	–0.042	0.075	–0.176	–0.175	–0.102	–0.307

TABLE 18 Spearman's rank-correlation coefficients between pre-selected candidate predictors from baseline (*continued*)

	Ankle catching	Plantar ROM flexion	Plantar ROM dorsiflexion	Pain at night	Difficulty with squatting	Difficulty with running
Days from injury to assessment	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Maximum bearable weight	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Sex	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Pain before injury	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Recurrent sprain	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Current employment	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Education	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Treatment arm	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
LTPA	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Walking	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Previous instability	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Previous instability frequency	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Injury mechanism	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Ankle grinding	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Ankle catching or locking	–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
Plantar ROM flexion	0.125	–	*ref.*	*ref.*	*ref.*	*ref.*
Plantar ROM dorsiflexion	0.090	0.667	–	*ref.*	*ref.*	*ref.*
Pain at night	0.175	0.169	0.119	–	*ref.*	*ref.*
Difficulty with squatting	0.087	0.177	0.164	0.130	–	*ref.*
Difficulty with running	0.081	0.074	0.061	0.053	0.441	–
Difficulty with jumping	0.081	0.074	0.061	0.053	0.441	1.000
Difficulty with twisting	0.040	0.086	0.083	0.091	0.475	0.859
Age	–0.032	0.024	–0.012	0.021	0.124	0.068
BMI	0.026	0.056	–0.076	0.059	0.006	–0.011
Able to bear weight	–0.109	0.131	0.127	–0.162	–0.018	–0.100
Pain when resting	0.215	0.193	0.106	0.434	0.138	–0.005
Pain when bearing weight	0.117	0.242	0.245	0.395	0.135	0.063
SF-12 mental component	–0.209	0.082	0.016	–0.088	0.000	–0.071

Difficulty with jumping	Difficulty with twisting	Age	BMI	Able to bear weight	Pain when resting	Pain when bearing weight
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
ref.	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
0.859	–	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
0.068	0.082	–	*ref.*	*ref.*	*ref.*	*ref.*
–0.011	0.048	0.227	–	*ref.*	*ref.*	*ref.*
–0.100	–0.131	0.054	–0.001	–	*ref.*	*ref.*
–0.005	0.047	0.065	0.158	–0.011	–	*ref.*
0.063	0.149	0.149	0.118	–0.069	0.630	–
–0.071	–0.063	–0.008	–0.126	0.238	–0.123	–0.128

TABLE 19 Summary of the full multivariable logistic regression models including all 23 candidate predictors of poor outcome after ankle sprain (outcomes 1 and 2)

	Outcome							
	1				2			
Variable	Coefficient	95% CI		*p*-value	Coefficient	95% CI		*p*-value
Age	**0.036**	**0.008**	**0.064**	**0.012**	0.015	–0.010	0.040	0.230
BMI	**0.039**	**–0.013**	**0.090**	**0.138**	0.012	–0.034	0.059	0.609
Pain when resting	**0.018**	**0.005**	**0.031**	**0.009**	**0.015**	**0.002**	**0.027**	**0.022**
Pain when bearing weight	**0.018**	**–0.001**	**0.037**	**0.057**	**0.013**	**–0.003**	**0.029**	**0.117**
SF-12 mental score	–0.006	–0.030	0.018	0.641	–0.012	–0.034	0.010	0.271
Sex (reference: male)								
Female	0.054	–0.581	0.689	0.868	–0.134	–0.734	0.466	0.661
Days from injury to assessment (reference: 0–2)								
≥ 3	**0.945**	**0.000**	**1.890**	**0.050**	**0.646**	**–0.129**	**1.421**	**0.101**
Able to bear weight at ED presentation (reference: no)								
Yes	0.538	–0.445	1.522	0.280	0.445	–0.376	1.266	0.285
Able to bear weight at baseline assessment (reference: no)								
Yes	**–0.848**	**–1.494**	**–0.202**	**0.010**	**–0.737**	**–1.328**	**–0.147**	**0.014**
Pain on the ankle before injury (reference: no)								
Yes	0.270	–0.499	1.038	0.491	0.120	–0.588	0.828	0.739
Recurrent sprain (reference: no)								
Yes	**1.355**	**0.486**	**2.224**	**0.002**	**1.207**	**0.396**	**2.018**	**0.004**
Pain in bed at night (reference: no)								
Yes	0.090	–0.572	0.752	0.790	–0.059	–0.647	0.528	0.843
Difficulty with squatting (reference: none/mild/moderate)								
Severe/extreme	–0.223	–0.976	0.531	0.561	0.005	–0.682	0.691	0.989
Current employment (reference: none)								
Part time	0.716	–0.163	1.595		0.452	–0.309	1.213	
Full time	0.685	–0.079	1.449	0.175	0.148	–0.517	0.813	0.500
Treatment received for ankle sprain (reference: tubular bandage)								
Below-knee cast	–0.554	–1.287	0.179		–0.504	–1.180	0.173	
Aircast brace	–0.394	–1.115	0.326		–0.451	–1.110	0.208	
Bledsoe boot	–0.218	–0.967	0.531	0.489	–0.442	–1.125	0.242	0.443
Education level (reference: CSE level or lower)								
O level/GCSE/A level	0.433	–0.432	1.298		**0.356**	**–0.443**	**1.154**	
Degree/higher degree	–0.217	–1.256	0.822	0.217	**–0.592**	**–1.542**	**0.358**	**0.042**

TABLE 19 Summary of the full multivariable logistic regression models including all 23 candidate predictors of poor outcome after ankle sprain (outcomes 1 and 2) (*continued*)

	Outcome							
	1				2			
Variable	Coefficient	95% CI		*p*-value	Coefficient	95% CI		*p*-value
Leisure-time physical activity (reference: none)								
< 1 time weekly	0.007	–1.198	1.211		0.301	–0.818	1.420	
> 1 time weekly	0.206	–1.055	1.466	0.794	0.263	–0.874	1.399	0.869
Walking 2 miles or more per day (reference: none)								
< 1 time weekly	–0.104	–0.867	0.659		–0.300	–1.000	0.401	
> 1 time weekly	–0.183	–0.811	0.444	0.847	–0.243	–0.788	0.303	0.626
Injury mechanism (reference: at home)								
Practising sports	0.115	–0.711	0.941		0.302	–0.457	1.062	
At work	0.444	–0.508	1.396		0.672	–0.206	1.550	
Outside, in public	–0.215	–0.966	0.535	0.524	–0.033	–0.727	0.662	0.323
Ankle grinding/clicking (reference: never)								
Rarely/sometimes	–0.226	–0.813	0.362		0.011	–0.531	0.553	
Often/always	–0.325	–1.245	0.596	0.696	0.048	–0.772	0.869	0.993
Ankle catching/locking (reference: never)								
Rarely/sometimes	0.224	–0.383	0.832		0.021	–0.525	0.568	
Often/always	0.487	–0.339	1.313	0.483	0.364	–0.362	1.090	0.602
Ankle ROM plantar flexion (reference: always/often)								
Sometimes/rarely	0.550	–0.380	1.479		0.474	–0.370	1.319	
Never	0.223	–0.826	1.273	0.395	–0.052	–1.002	0.897	0.185
Ankle ROM dorsiflexion (reference: always/often)								
Sometimes/rarely	–0.019	–1.041	1.002		–0.127	–1.017	0.762	
Never	0.366	–0.734	1.466	0.528	0.418	–0.568	1.404	0.253
Intercept	**–3.003**	**–5.162**	**–0.845**	**0.007**	**–2.045**	**–3.892**	**–0.198**	**0.030**

A level, Advanced level; CSE, Certificate of Secondary Education; GCSE, General Certificate of Secondary Education; O level, Ordinary level; ROM, range of motion; SF-12, Short Form questionnaire-12 items.

Notes

Candidate predictors that were statistically significant in accordance with AIC are in bold. Although education level was a statistically significant predictor for outcome 2, a decision has been made not to include it in the final model.

TABLE 20 Estimates of the final models for the prediction of outcomes 1 and 2 occurrence

Variable	Outcome					
	1			2		
Baseline models	**Coefficient**	**95% CI**	***p*-value**	**Coefficient**	**95% CI**	***p*-value**
Age	0.027	0.006 to 0.048	0.014	–	–	–
BMI	0.031	–0.014 to 0.076	0.178	–	–	–
Pain when resting	0.016	0.005 to 0.027	0.005	0.014	0.004 to 0.024	0.008
Pain when bearing weight	0.019	0.004 to 0.035	0.016	0.015	0.001 to 0.029	0.033
Days from injury to assessment (reference: 0–2 days)						
≥ 3	0.854	0.068 to 1.640	0.034	0.650	0.019 to 1.280	0.043
Able to bear weight at baseline (reference: no)						
Yes	–0.792	–1.376 to –0.207	0.008	–0.705	–1.225 to –0.184	0.008
Recurrent sprain (reference: no)						
Yes	1.180	0.417 to 1.944	0.003	1.100	0.388 to 1.813	0.003
Intercept	–1.580	–2.152 to –1.008	< 0.001	–1.080	–1.513 to –0.647	< 0.001
Updated models (baseline + week 4 predictors)	**Coefficient**	**95% CI**	***p*-value**	**Coefficient**	**95% CI**	***p*-value**
Age	0.018	–0.005 to 0.040	0.127	–	–	–
BMI	0.025	–0.022 to 0.072	0.292	–	–	–
Pain when resting	0.010	–0.002 to 0.022	0.107	0.005	–0.006 to 0.016	0.381
Pain when bearing weight	0.014	–0.002 to 0.030	0.092	0.010	–0.004 to 0.024	0.176
Pain when bearing weight 4 weeks after injury	0.022	0.012 to 0.032	< 0.001	0.026	0.016 to 0.035	< 0.001
Days from injury to assessment (reference: 0–2 days)						
≥ 3	0.702	–0.117 to 1.520	0.092	0.444	–0.230 to 1.118	0.194
Able to bear weight at baseline (reference: no)						
Yes	–0.802	–1.412 to –0.192	0.010	–0.741	–1.288 to –0.194	0.008
Recurrent sprain (reference: no)						
Yes	1.170	0.386 to 1.953	0.004	1.168	0.416 to 1.919	0.002
Intercept	–1.543	–2.128 to –0.958	< 0.001	–1.012	–1.468 to –0.557	< 0.001

Note

Linear terms selected by the MFP for continuous predictors for both outcomes 1 and 2: age, 29.88 years; BMI, 26.32 kg/m^2; pain when resting, 37.75 points; pain when bearing weight, 75.40 points; pain when bearing weight at 4 weeks after injury, 36.23 points.

TABLE 21 Summary of the full multivariable logistic regression including the predictors selected for the baseline models and the 4-weeks candidate predictors for the updated models

	Outcome					
	1			2		
Variable	Coefficient	95% CI	*p*-value	Coefficient	95% CI	*p*-value
Baseline predictors						
Age	0.020	–0.004 to 0.044	0.097	–	–	–
BMI	0.024	–0.027 to 0.074	0.356	–	–	–
Pain when resting	0.008	–0.005 to 0.021	0.228	0.004	–0.009 to 0.016	0.554
Pain when bearing weight	0.014	–0.002 to 0.031	0.090	0.010	–0.005 to 0.024	0.199
Days from injury to assessment (reference: 0–2 days)						
≥ 3	0.639	–0.288 to 1.565	0.174	0.450	–0.302 to 1.202	0.238
Able to bear weight at baseline assessment (reference: no)						
Yes	–0.877	–1.531 to –0.223	0.009	–0.797	–1.380 to –0.214	0.007
Recurrent sprain (reference: no)						
Yes	1.158	0.306 to 2.009	0.008	1.148	0.378 to 1.918	0.004
Week 4 predictors						
Pain when bearing weight 4 weeks after injury	**0.019**	**0.005 to 0.033**	**0.007**	**0.026**	**0.013 to 0.039**	**< 0.001**
Another injury (reference: no)						
Yes	–0.387	–1.454 to 0.680	0.476	0.254	–0.642 to 1.151	0.577
Returned to sports activities (reference: yes)						
No	–0.173	–0.785 to 0.440	0.580	–0.093	–0.636 to 0.449	0.736
Difficulty with running, jumping or twisting (pivoting) 4 weeks after injury (reference: no)						
Yes	0.041	–0.801 to 0.882	0.924	–0.420	–1.139 to 0.299	0.252
Pain in bed at night 4 weeks after injury (reference: no)						
Yes	0.555	–0.453 to 1.563	0.279	0.489	–0.481 to 1.459	0.322
Difficulty with squatting 4 weeks after injury (reference: no)						
Yes	0.137	–0.603 to 0.877	0.716	0.361	–0.366 to 1.088	0.329
Ankle swelling 4 weeks after injury (reference: never)						
Rarely/sometimes	0.692	–0.391 to 1.775		0.656	–0.308 to 1.619	
Often/always	0.427	–0.737 to 1.590	0.384	0.523	–0.501 to 1.546	0.404
Ankle grinding/clicking (reference: never)						
Rarely/sometimes	0.652	–0.036 to 1.340		0.608	0.010 to 1.206	
Often/always	0.409	–0.480 to 1.298	0.177	0.275	–0.515 to 1.066	0.313
Ankle catching/locking (reference: never)						
Rarely/sometimes	–0.279	–0.982 to 0.423		–0.372	–1.004 to 0.260	
Often/always	0.622	–0.410 to 1.654	0.193	0.738	–0.261 to 1.738	0.497

continued

TABLE 21 Summary of the full multivariable logistic regression including the predictors selected for the baseline models and the 4-weeks candidate predictors for the updated models (*continued*)

	Outcome					
	1			2		
Variable	Coefficient	95% CI	*p*-value	Coefficient	95% CI	*p*-value
Ankle ROM dorsiflexion 4 weeks after injury (reference: always/often)						
Sometimes/rarely	–0.387	–1.055 to 0.281		–0.229	–0.827 to 0.368	
Never	0.500	–0.512 to 1.513	0.159	0.408	–0.543 to 1.359	0.387
Intercept	–2.215	–3.433 to –0.997	< 0.001	–1.594	–2.674 to –0.515	0.004

ROM, range of motion.
Notes
The bold signifies candidate predictors that were statistically significant in accordance with AIC criteria. Although education level was a statistically significant predictor for outcome 2, a decision has been made not to include it in the final model.

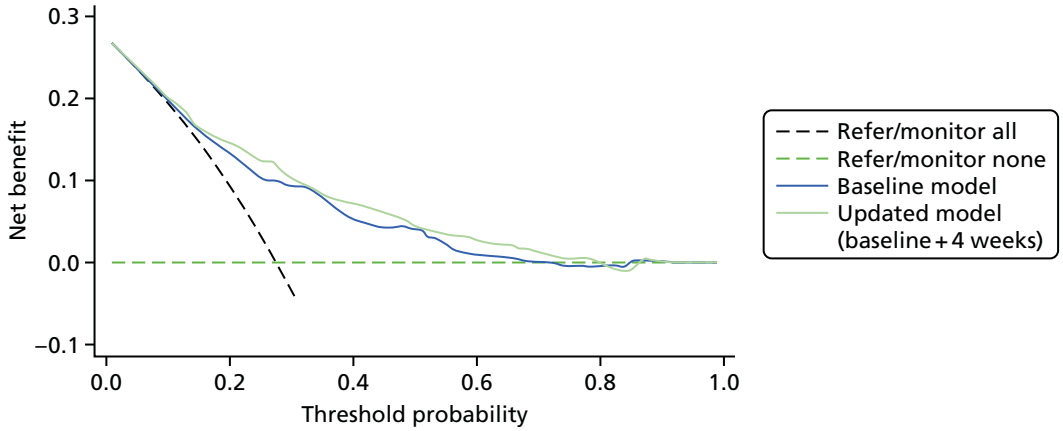


FIGURE 11 Decision curve analysis plot for outcome 1. Adapted with permission from Schlüssel *et al.*[26] © Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY. Published by BMJ. This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

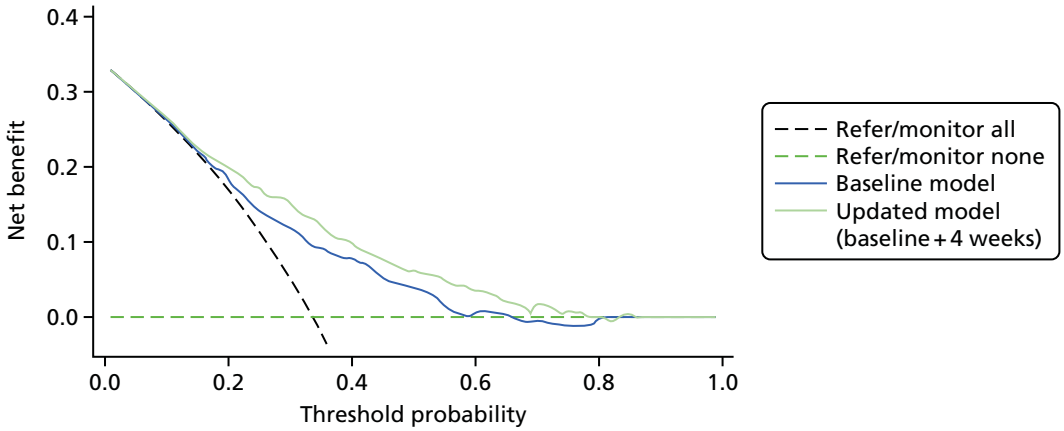


FIGURE 12 Decision curve analysis plot for outcome 2.

Calibration plots overlying the results of the analysis on the 50 imputed data sets are presented in *Figures 13* and *14*. Perfect predictions should lie on the 45° line in the calibration plot for agreement with the outcome. As anticipated, on average, the calibration across all models was consistently strong, with close agreement between the observed and predicted risks of developing outcomes 1 (*Figure 13*) and 2 (*Figure 14*). Shrinkage suggested both prognostic models to be unstable, with a considerable amount of optimism. The heuristic shrinkage factor for the coefficients of the predictors in the baseline prognostic model for outcome 1 was 0.71, suggesting that 29% of the model fit was non-replicable noise. For the updated versions (baseline and week 4 predictors) of both prognostic models, the estimated heuristic shrinkage factor was 0.84. The shrunk coefficients and intercepts are presented in *Table 22*.

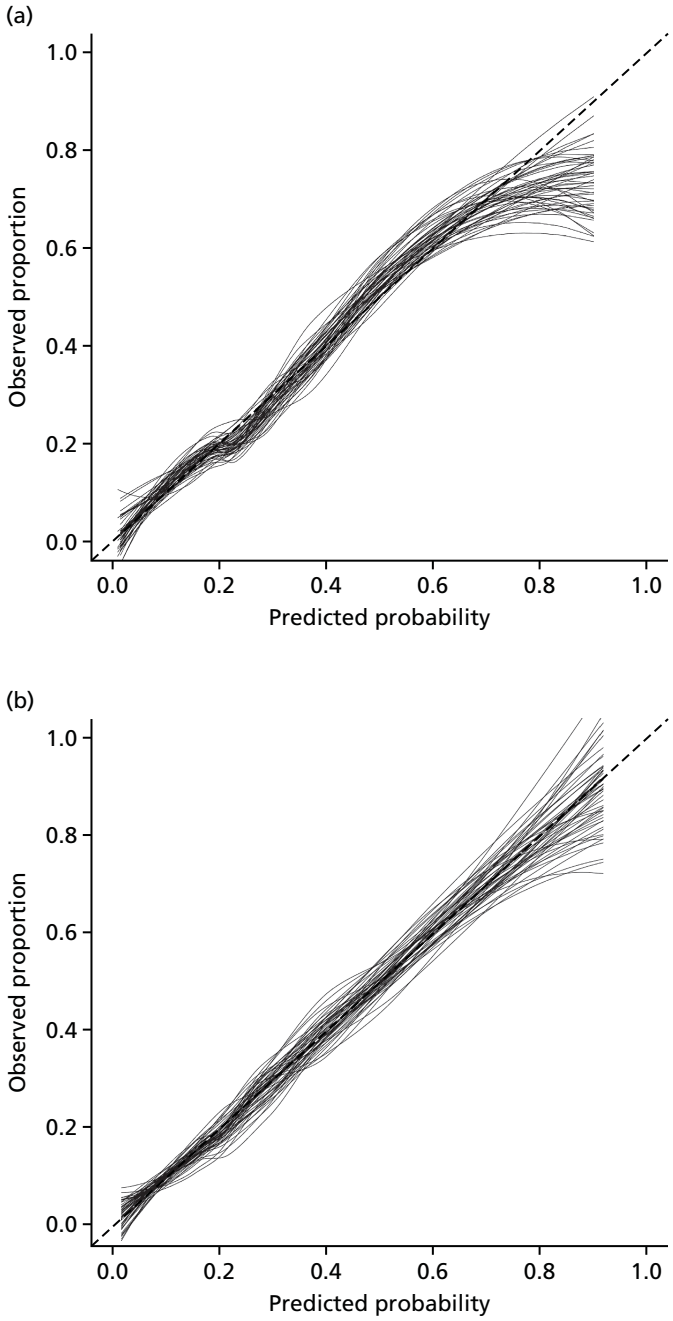


FIGURE 13 Calibration plot of the prognostic model for outcome 1. (a) The baseline model; and (b) the updated (baseline plus 4 weeks) model.

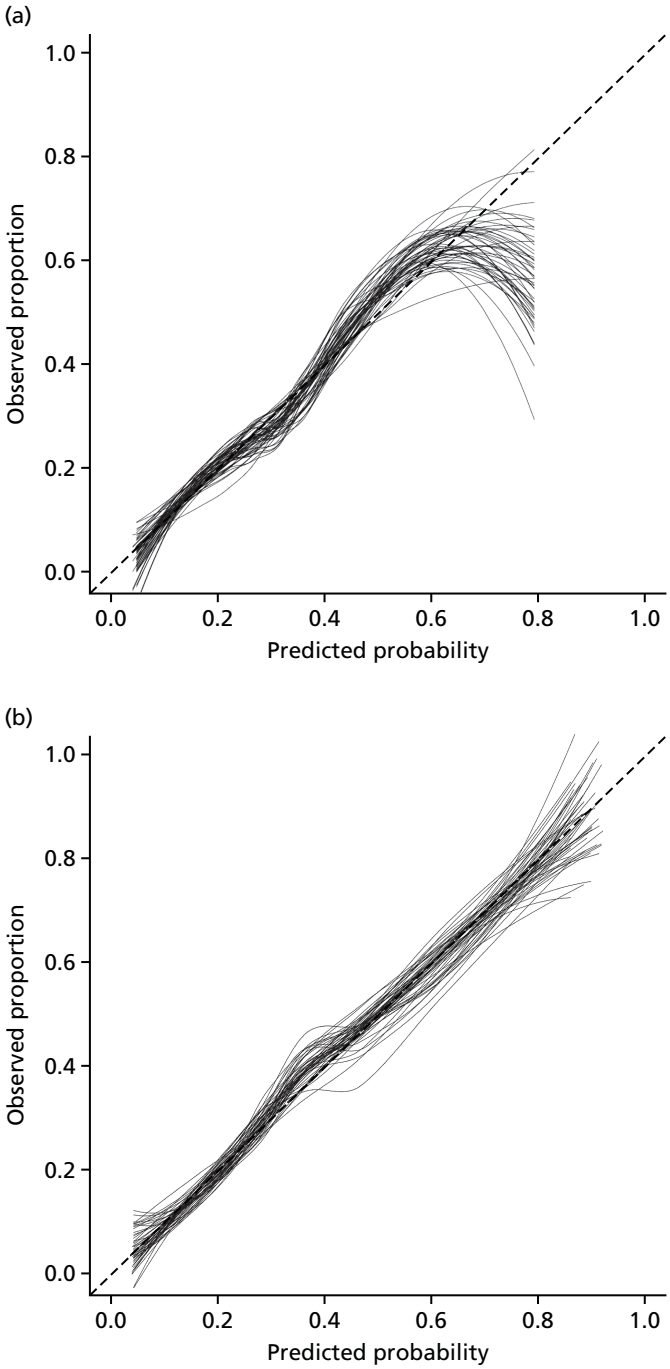


FIGURE 14 Calibration plot of the prognostic model for outcome 2. (a) The baseline model; and (b) the updated (baseline plus 4 weeks) model.

TABLE 22 Intercept and regression coefficients of the prediction models for poor recovery 9 months after ankle sprain (outcomes 1 and 2), before and after correction for optimism (shrinkage)

	Outcome			
	1		2	
Predictors in the baseline models	Coefficient	Shrunk coefficient	Coefficient	Shrunk coefficient
Age	0.027	0.019	–	–
BMI	0.031	0.022	–	–
Pain when resting	0.016	0.011	0.014	0.008
Pain when bearing weight	0.019	0.014	0.015	0.009
> 2 days from injury to assessment	0.854	0.605	0.650	0.396
Able to bear weight on the injured ankle	–0.792	–0.561	–0.705	–0.429
Recurrent sprain	1.180	0.836	1.100	0.670
Intercept	–1.580	–1.363	–1.080	–0.903
Predictors in the updated models (baseline + 4-week variables)				
Age	0.018	0.015	–	–
BMI	0.025	0.021	–	–
Pain when resting	0.010	0.008	0.010	0.010
Pain when bearing weight	0.014	0.012	0.010	0.010
Pain when bearing weight 4 weeks after injury	0.022	0.018	0.026	0.022
> 2 days from injury to assessment	0.702	0.591	0.444	0.373
Able to bear weight on the injured ankle	–0.802	–0.676	–0.741	–0.623
Recurrent sprain	1.170	0.985	1.168	0.982
Intercept	–1.543	–1.420	–1.012	–0.942

Application of the SPRAINED study model

The following section will provide an example of how the internally validated SPRAINED study model can be applied in practice. To make predictions with the SPRAINED prognostic models, the following equations are required (please note that all linear terms selected by the MFP for continuous predictors were incorporated into the models' intercepts).

Baseline model for outcome 1

$$\begin{aligned} Y = &-3.68 + (0.02 \times \text{age}) + (0.02 \times \text{BMI}) + (0.01 \times \text{pain when resting}) \\ &+ (0.01 \times \text{pain when bearing weight}) + (0.61 \text{ if days from injury to assessment} > 2) \\ &- (0.56 \text{ if able to bear any weight on the injured ankle}) \\ &+ (0.84 \text{ if the injury is a recurrent sprain}). \end{aligned} \quad (2)$$

Then, we need to convert the log-odds (Y) into probability. This can be done by applying the following equation:

$$P = 1/[1 + \exp(-Y)], \quad (3)$$

where P is the probability of developing the outcome and Y is the log-odds estimated with the model.
To provide a practical example of how to use the SPRAINED prognostic model to predict the occurrence of outcome 1, we consider a hypothetical patient:

> *Patient with ankle sprain, male, 38 years old, presenting at the ED 3 days after occurrence of the injury, with an estimated BMI of 25.6 kg/m^2, reporting pain when resting of 50 points on the visual analogue scale (VAS), and 80 [points] when bearing some weight on the injured ankle, willing to bear weight on the ankle and stating that this is a recurrent injury, attributable to the practice of basketball.*

To calculate the risk of having a poor recovery from this ankle sprain 9 months after the injury, the information on the relevant predictors must be entered in the model shown in *Table 23*.

Applying *Equation 2*:

(1) $Y = -3.68 + (0.02 \times 38) + (0.02 \times 25.6) + (0.01 \times 50) + (0.01 \times 80) + 0.61 - 0.56 + 0.84$.
(2) $Y = -3.68 + 0.76 + 0.51 + 0.50 + 0.80 + 0.61 - 0.56 + 0.84$.
(3) $Y = -0.22$.

Applying the transformation (*Equation 3*):

(4) $P = 1/[1 + \exp(0.22)]$.
(5) $P = 1/(1 + 1.24)$.
(6) $P = 1/2.24$.
(7) $P = 0.45$ (or 45%).

The estimated probability of a poor outcome developing 9 months after ankle sprain (as per the definition of outcome 1) for that patient would be 45%.

If we had the chance of reassessing the patient 4 weeks after the injury, assessed their pain when bearing weight at this stage (for example, 30 on a scale from 0 to 100) and applied the updated model (baseline + 4-week predictors), the following would need to be done.

TABLE 23 Information on the relevant prediction models for poor recovery 9 months after ankle sprain

Predictor	Information
Age (years)	38
BMI (kg/m^2)	25.6
Pain when resting (VAS score, points)	50
Pain when bearing weight (VAS score, points)	80
> 2 days from injury to assessment	Yes
Able to bear weight on the injured ankle	Yes
Recurrent sprain	Yes

VAS, visual analogue scale.

Updated model for outcome 1 (baseline + 4-week predictors)

$$\begin{aligned} Y = &-4.4 + (0.01 \times \text{age}) + (0.02 \times \text{BMI}) + (0.01 \times \text{pain when resting}) \\ &+ (0.01 \times \text{pain when bearing weight}) + (0.59 \text{ if days from injury to assessment} > 2) \\ &- (0.68 \text{ if able to bear any weight on the injured ankle}) \\ &+ (0.99 \text{ if the injury is a recurrent sprain}) \\ &+ (0.02 \times \text{pain when bearing weight 4 weeks after injury}). \end{aligned} \quad (4)$$

Applying *Equation 4*:

(1) $Y = -4.4 + (0.01 \times 38) + (0.02 \times 25.6) + (0.01 \times 50) + (0.01 \times 80) + 0.59 - 0.68 + 0.99 + (0.02 \times 30)$
(2) $Y = -4.4 + 0.38 + 0.51 + 0.50 + 0.80 + 0.59 - 0.68 + 0.99 + 0.60$
(3) $Y = -0.71$.

Applying the transformation (*Equation 3*):

(4) $P = 1/[1 + \exp(0.71)]$
(5) $P = 1/(1 + 2.03)$
(6) $P = 1/3.03$
(7) $P = 0.33$ (or 33%).

Therefore, by adding extra information on the patient follow-up, we were able to estimate a more precise probability of presenting with poor outcome at 9 months after injury.

To calculate the risk of having poor recovery at 9 months after ankle sprain according to the definition of outcome 2, the following model should be applied.

Baseline model for outcome 2

$$\begin{aligned} Y = &-2.07 + (0.01 \times \text{pain when resting}) + (0.01 \times \text{pain when bearing weight}) \\ &+ (0.40 \text{ if days from injury to assessment} > 2) \\ &- (0.43 \text{ if able to bear any weight on the injured ankle}) \\ &+ (0.67 \text{ if the injury is a recurrent sprain}). \end{aligned} \quad (5)$$

Applying *Equation 5*:

(1) $Y = -2.07 + (0.01 \times 50) + (0.01 \times 80) + 0.40 - 0.43 + 0.67$
(2) $Y = -2.07 + 0.50 + 0.80 + 0.40 - 0.43 + 0.67$
(3) $Y = -0.13$.

Applying the transformation (*Equation 3*):

(4) $P = 1/[1 + \exp(0.14)]$
(5) $P = 1/(1 + 1.14)$
(6) $P = 1/2.14$
(7) $P = 0.47$ (or 47%).

For the same patient, the probability of a poor outcome developing 9 months after ankle sprain (as per outcome 2 definition) would be slightly higher (47%) than that obtained when using the outcome 1 definition.

To calculate the updated probability of this patient presenting poor outcome at 9 months using the model with baseline and 4 weeks predictors (considering that in the reassessment, their pain score when bearing weight was 30), the following equation should be applied.

Updated model for outcome 2 (baseline + 4-week predictors)

$$\begin{aligned} Y = &-2.79 + (0.01 \times \text{pain when resting}) + (0.01 \times \text{pain when bearing weight}) \\ &+ (0.37 \text{ if days from injury to assessment} > 2) \\ &- (0.62 \text{ if able to bear any weight on the injured ankle}) \\ &+ (0.98 \text{ if the injury is a recurrent sprain}) \\ &+ (0.02 \times \text{pain when bearing weight 4 weeks after injury}). \end{aligned} \quad (6)$$

Applying *Equation 6*:

(1) $Y = -2.79 + (0.01 \times 50) + (0.01 \times 80) + 0.37 - 0.62 + 0.98 + (0.02 \times 30)$
(2) $Y = -2.79 + 0.50 + 0.80 + 0.37 - 0.62 + 0.98 + 0.60$
(3) $Y = -0.16$.

Applying the transformation (*Equation 3*):

(4) $P = 1/[1 + \exp(0.16)]$
(5) $P = 1/(1 + 1.17)$
(6) $P = 1/2.17$
(7) $P = 0.46$ (or 46%).

Therefore, by adding extra information on the patient follow-up, the updated probability of presenting with poor outcome at 9 months after injury was 46%.

The observational cohort study, conducted to enable external validation of the prognostic models presented, is reported in the following chapter. The results of the prognostic model development and external validation are summarised and discussed together in *Chapter 7*.

Chapter 6 External validation study of the SPRAINED prognostic models

Introduction

This chapter describes the external validation process of the two prognostic models (and their updates) developed to predict the risk of poor outcome at 9 months after an acute ankle sprain. A prospective observational cohort study was conducted with the aim of obtaining data to externally validate and optimise the prognostic models for use in EDs. Before participant recruitment began, the models were developed, corrected for optimism and updated with the inclusion of an additional predictor for which information was collected at 4 weeks after the injury using the CAST data set (see *Chapter 5*), which was subsequent to a systematic literature review (see *Chapter 3*) and a consensus process involving clinician and patient perspectives (see *Chapter 4*).

Methods

Cohort design and study population

People with acute ankle sprain attending 10 NHS EDs across England were recruited for the SPRAINED cohort study (see *Acknowledgements* for details on recruiting centres) over a period of 9 months (July 2015–March 2016). This was an observational cohort study; therefore, participants were not randomised, nor did they receive any interventions other than usual care at each site. Data collection took place at the time of a participant's presentation to any of the study recruiting sites (baseline) and subsequently at 4 weeks and 4 and 9 months after the initial injury.

People were invited to take part in the study if they met the following inclusion criteria and none of the exclusion criteria.

Inclusion criteria

- Participant was willing and able to give informed consent for participation in the study.
- Aged ≥ 16 years.
- Diagnosed with acute ankle sprain (grades I to III, < 7 days old).

Exclusion criteria

- Ankle fracture (apart from flake fractures < 2 mm).
- Other recent (< 3 months) lower limb fracture.

Sample size

The recommended sample size estimation for an external validation of a prognostic model is that 100 outcome events are required, this being the minimum number needed to ensure accurate estimation of the calibration of the model.[72,76] The event rates for the outcomes of interest in CAST were between 26% and 32%, depending on the definition of the outcomes (three symptoms and four symptoms/clinical events, respectively), this would require an overall sample size of between 313 and 385 participants. Assuming a rate of 25% for loss to follow-up and a lower event rate (20%) when recruiting all grades of ankles sprains, a minimum of 675 participants were targeted for recruitment to increase the chances of achieving the required event rates. We anticipated recruiting people with a range of sprains, including grades I to III.

Screening and eligibility assessment

People were screened by clinicians on admission to EDs and assessed for eligibility to take part in the SPRAINED cohort study. A member of the research team at the study centres administered the study clinical data set form (CDF) and recorded responses and findings from the clinical examination (see *Appendix 4*). The short CDF served three purposes:

1. collection of routine core clinical data set in a tick-box format (reflecting the data that would be normally recorded in the course of routine clinical practice)
2. to record, via a tick box, that clinicians had provided potential participants with the trial information pack and a brief explanation of the trial
3. to record, via a tick box, whether or not the individual had given permission for a member of the research team to make contact with him/her to discuss the study further and complete the informed consent process.

One copy of the CDF was filed in the person's medical notes as a treatment record and a second copy, when agreement was given, was passed to the local research team. The team member then contacted the individual and continued the informed consent process. Only once consent was obtained was the clinical data set sent to the central study office. The clinical data set of any person who did not agree to study participation remained at the site in his/her medical notes.

Informed consent and recruitment

The initial approach was made by a member of the ED clinical team. A verbal explanation of the study, along with a study information leaflet, was given to all potentially eligible people. Posters were displayed in all participating departments to inform participants that the study was occurring.

The informed consent process was carried out by a registered health-care professional with delegated authority from the principal investigator at the recruiting site. Before consenting to participate in the study, the person was asked by a member of the local clinical team for permission to allow the local research team to speak to them, either in person or by telephone, to take forward the informed consent process. Formal consent to participation was provided either in person, by post or by telephone. Before any data were provided to the study team, the participant personally signed and dated the latest approved version of the informed consent form (ICF), or verbal consent was recorded by a member of the local team on a form during the informed consent telephone call. The participant had the opportunity to question the clinical/research team, and to consult their general practitioner (GP) or other independent parties to decide whether or not they would participate in the study.

Written informed consent was obtained by means of participant-dated signature and dated signature of the person who presented and obtained the informed consent. Verbal informed consent was obtained by means of the dated signature of the local team member taking consent over the telephone. A copy of the completed written or verbal ICF was retained by the participant (or posted to the participant in the case of oral consent). One copy was sent to the study co-ordinating team in Oxford. The original signed consent form was retained in the medical notes, and a copy was held in the investigator site file.

Participants consented to allow the study team to use the CDF completed during the ED attendance and an additional questionnaire 4 weeks after this (SPRAINED prognostic model and any additional important information), as well as follow-up questionnaires at 4 and 9 months, which aimed to map the recovery trajectory and final recovery status at 9 months. A questionnaire at 4 months served as a reminder of the study and, as loss to follow-up was likely to increase over time, helped to ensure that responses on the core components of the outcomes of interest were available for as many participants as possible.

Data collection and management

Baseline data were collected from participants and recorded on a paper CDF. Data for the three study follow-up points [4 weeks (prognostic variables) and 4 and 9 months (outcome data) after baseline

assessment] were collected by using paper case report forms (CRFs) sent to participants via post, or completed by telephone call when necessary. The telephone calls enabled collection of at least the core data on the outcome measures for participants that did not return the questionnaire to the trial office. When preferred by the participant, secure online data collection took place for the 4-week time frame.

Baseline CDFs were sent by a member of the local research team to the study co-ordinating office in Oxford by post. Follow-up CRFs were sent by the participant to the study co-ordinating office in Oxford by post, using a Freepost return envelope. When telephone follow-up was used, a member of the central study team recorded data directly onto the relevant forms.

On receipt of data forms (CDFs and CRFs), appropriate data quality and validation checks were carried out and the data were entered into a study-dedicated database, which was developed and maintained by OCTRU, a UK Clinical Research Network (UKCRN)-registered clinical trials unit. OpenClinica software (OpenClinica LLC, Waltham, MA, USA) was used to develop and maintain the study database. To identify manual entry errors, a 10% double-entry check was carried out at regular intervals during the data collection phase of the study.

Details relating to ethics approvals and monitoring are outlined in *Ethics approval and monitoring*.

Study assessments

Baseline assessments

Baseline data were collected on the clinician-completed CDF and included:

- demographics (name, age, contact details)
- patient history
- clinical examination
- clinical investigation
- clinical management
- clinical diagnosis
- prognostic factors
- agreement for research team to contact patient.

Participant contact details were also collected at baseline to facilitate study follow-up. This included full name, address, NHS number, mobile and/or telephone number, e-mail address and a preferred time to be contacted. Reasons for declining the study were collected, if given.

Follow-up assessment 1 (prognostic variables at 4 weeks after ankle sprain)

Follow-up at 4 weeks after ankle sprain was conducted by electronic, telephone or postal questionnaire. Questions included:

- current clinical status (recurrence of injury, swelling or pain in the ankle)
- return to normal activities.

Follow-up assessments 2 and 3 (outcome variables at 4 and 9 months after ankle sprain)

Follow-up at 4 and 9 months after ankle sprain was conducted by postal or telephone questionnaire. Information elicited included:

1. recurrence of injury
2. FAOS
3. health service resource use
4. health-related quality of life [EuroQol-5 Dimensions (EQ-5D)].

Outcome measures

For the external validation data set (SPRAINED observational cohort study), poor outcome at 9 months after ankle sprain was defined in the same way as it was in the development study (see *Chapter 5*). The same questions were asked to SPRAINED study participants, so the same two outcomes could be constituted. Therefore, the definition of poor outcome was the presence of any, or a combination, of the following symptoms or clinical events (for further details see *Definition of the primary outcomes*).

Outcome 1

- Severe persistent pain.
- Severe functional difficulty.
- Significant lack of confidence in the ankle.

Outcome 2

- Severe persistent pain.
- Severe functional difficulty.
- Significant lack of confidence.
- Recurrent sprain.

Predictors of poor outcome at 9 months after ankle sprain

All variables included in the prognostic models developed to predict the occurrence of poor outcome at 9 months after ankle sprain (the SPRAINED prognostic models, see *Chapter 5*) were included in the baseline CRFs. Data collection on a few additional candidate predictors that were not included in the final models was also conducted to allow some room for model updating, if necessary. However, the data collected at baseline were kept to a minimum, prioritising the predictors included in the two developed models and those candidate predictors that the consensus group considered to have the most clinical importance and relevance to patients. Except for pain scores (collected as discrete variables in the SPRAINED cohort study), data collection on all variables was performed respecting their original format in the CAST data set. A complete list of the variables collected at baseline in both CAST and the SPRAINED cohort study, with formats and number of missing data, is given in *Table 24*.

Statistical methods

Exploratory analysis and data transformation

Baseline characteristics of participants were summarised using means, SDs and ranges for continuous variables, or counts and percentages for categorical variables. To examine differences in case mix between the participants in the development (CAST) and external validation (SPRAINED cohort study), characteristics of participants included in the two studies were compared narratively (no statistical tests were performed).

Categorical variables were recategorised by collapsing some of their categories, to match the format of those included in the regression analyses during the model development stage. The distribution of the continuous predictors was also assessed, first considering their empirical distributions by producing histograms and then by assessing these for normality by means of normal probability plots, box plots and dot plots. The presence of any outliers was assessed based on visual examination of the box plots. Extreme values were inspected to confirm whether or not they were clinically plausible.

Handling missing data

As there was more than one predictor with missing data in the SPRAINED observational cohort study that was needed to validate the model (up to 8%, for BMI), MICE was used to replace missing values (see *Tables 22* and *26* for percentages of missing data for predictor variables and outcomes, respectively). MICE uses a set of imputation equations, including one for each of the predictors with missing data; all equations include all of the predictors included in the prediction model, predictors of predictors and the

TABLE 24 Predictor variables and their formats in the original CAST data set, the final model and the SPRAINED data set (with respective numbers of missing data)

	CAST data set		Modelling process/final model		SPRAINED data set		
Variable	Type	Categories/units	Type	Categories/units	Type	Categories/units	Missing (%)
Sex	Binary	• Male • Female	Binary	• Male • Female	Binary	• Male • Female	–
Recurrent sprain[a]	Binary	• Yes • No	Binary	• No • Yes	Binary	• No • Yes	6.5
Able to bear weight on the injured ankle	Continuous	kg	Binary	• No • Yes[b]	Binary	• No • Yes	0.7
Employment status	Categorical	• None • Part time • Full time • Student • Retired	Categorical	• None[c] • Part time • Full time	Categorical	• No • Part time • Full time • Student • Retired	0.3
Injury setting	Categorical	• During sport • At work • At home • In public • Other	Categorical	• During sport • At work • At home • In public/other	Categorical	• During sport • At work • At home • In public • Other	2.1
Ankle/foot catching/locking	Categorical	• Never • Rarely • Sometimes • Often • Always	Categorical	• Never • Rarely • Sometimes • Often • Always	Categorical	• Never • Rarely • Sometimes • Often • Always	3.1
Ankle ROM plantar flexion	Categorical	• Always • Often • Sometimes • Rarely • Never	Categorical	• Often • Always • Rarely • Sometimes • Never	Categorical	• Never • Rarely • Sometimes • Often • Always	1.5

continued

TABLE 24 Predictor variables and their formats in the original CAST data set, the final model and the SPRAINED data set (with respective numbers of missing data) (*continued*)

	CAST data set		Modelling process/final model		SPRAINED data set		
Variable	Type	Categories/units	Type	Categories/units	Type	Categories/units	Missing (%)
Ankle ROM dorsiflexion	Categorical	• Always • Often • Sometimes • Rarely • Never	Categorical	• Often • Always • Rarely • Sometimes • Never	Categorical	• Never • Rarely • Sometimes • Often • Always	1.6
Age	Continuous	Years	Continuous	Years	Continuous	Years	–
Days from injury to assessment[d]	Continuous	Days	Binary	• 1–2 days • 3–7 days	Continuous	Days	–
BMI[e]	Continuous	kg/m^2	Continuous	kg/m^2	Continuous	kg/m^2	8.2
Pain at rest	Continuous	0–100	Continuous	0–100	Discrete	0–10	3.4
Pain at weight bearing	Continuous	0–100	Continuous	0–100	Discrete	0–10	4.4
Pain at weight bearing at 4 weeks	Continuous	0–100	Continuous	0–100	Discrete	0–100	50

ROM, range of motion.
a An ankle sprain that has happened to a previously injured ankle (at least twice), with the last injury occurring in the previous 12 months.
b Any value different from 0.
c Combination of unemployed, student and retired.
d Not allowed > 7 days in the CAST data set.
e Calculated from height and weight (both continuous variables), as collected in the baseline CAST questionnaire.

outcomes. It is recommended that the imputation models should take into account all predictors within the analysis model as well as the outcome (to be predicted by the prognostic model). Including more predictors within the imputation model makes the MAR assumption more plausible by potentially including factors that may explain the missingness. Multiple imputation was performed, assuming that all missing variable data were MAR. This missing data mechanism assumes that the probability of an observation being missing is dependent on the observed data. To reflect the uncertainty in the imputation, 50 imputed data sets were created. The models were independently estimated for outcomes 1 and 2, and imputations were therefore performed in separate procedures, producing two different sets of 50 complete data sets (see *Chapter 5* for more details on the MICE principles, structure and commands used when handling missing data). Each of the imputed data sets was analysed separately by calculating the model discrimination and calibration. Combined calibration plots overlaying the calibration lines of the 50 analysed data sets for each outcome were produced. Discrimination is also presented for each model, in terms of *c*-statistics combined across the 50 analysed data sets for each outcome using Rubin's rules.[67]

Model performance

The performance of the prognostic models was assessed in terms of calibration and discrimination. Calibration was defined as follows: 'for patients with a predicted risk of *R*%, on average *R* out of 100 should indeed suffer from the disease or event of interest'. Calibration was assessed graphically by plotting the observed outcomes (on the *y*-axis) against the predicted probabilities from the models (on the *x*-axis). To produce the plots, participants were ranked from lowest predicted risk to highest predicted risk and grouped into tenths of predicted risk (i.e. 10 equal-sized groups). For each of the 10 groups, the mean predicted risk and the proportion of observed outcomes were calculated and plotted against each other. A flexible calibration curve was also fit using LOWESS to capture the agreement (and any miscalibration) between the observed outcomes and predicted probabilities over the entire probability range.[72]

Discrimination reflects the ability of the model to distinguish between participants who do and those who do not experience an event during the study period. Discrimination was assessed using the *c*-statistic, where a value of 0.5 represents chance and 1 represents perfect discrimination.[77] The *c*-statistic was classified as follows: 0.5–0.6, fail; 0.6–0.7, poor; 0.7–0.8, fair; 0.8–0.9, good; and 0.9–1.0, excellent. Individual probabilities of developing the outcomes were estimated by applying the developed prognostic models to each participant in the SPRAINED observational cohort study data set. Model performance was assessed for both the baseline and updated (baseline + 4-week predictors) models.

Finally, to estimate the benefit of using the developed prognostic models, the probabilities of developing poor outcome were estimated using the models' equations and participants were ranked on the basis of their estimated risks. These probabilities were used to calculate the number of people per 1000 identified as being at high risk of a poor outcome, according to different selected thresholds, and how many of these people go on to present with one of the outcomes compared with a strategy in which all individuals are deemed at high risk of a poor outcome.

Subgroup analysis

The rate of poor outcome at 9 months in the SPRAINED data set was expected to be lower than the rate observed in the CAST data set. One of the inclusion criteria for CAST stated that patients would be included if they had been diagnosed with an ankle injury of grade 2 (moderate severity) or 3 (severe), and so were more likely to have a poor outcome. In the SPRAINED cohort study, presenting with an injury of grade 1 (mild severity) was not an exclusion criterion, as the aim was to recruit a more representative sample of the population with this type of injury seeking medical assistance in the NHS. Therefore, a subgroup analysis was performed, with the aim of applying the prognostic models to a subsample of individuals composed of those presenting with injury severity of grades 2 or 3 (more similar to the population in the development data set), to check whether or not the models would present better performance among this specific group of patients. Model performance in the subgroup of patients with moderate or severe injuries was assessed for both the baseline and updated (baseline + 4-week predictors) models.

Model recalibration

In case of poor performance of the developed models, a strategy of recalibrating the models was planned. Recalibration methods may include adjustment of the intercept, additional adjustment of predictor coefficients (using the same method adopted during the development phase or a different approach), re-estimating predictor coefficients, and adding or removing predictors from the original model.[78] The adopted approach was to re-estimate the intercepts and predictor coefficients (refit the model in the SPRAINED observation cohort study data set). The prognostic models were refitted using a logistic regression modelling framework with the logit probability of an adverse outcome as the response variable. The same predictors selected for the two prognostic models were included together in full logistic regression models as independent variables and no exclusion based on the statistical significance of their adjusted relationship with the outcomes was made. Continuous variables were kept as continuous to avoid loss of prognostic information; the shape of the relationships between continuous predictors and the outcome were investigated and modelling performed using the MFP algorithm when appropriate. The 'best transformation' for each continuous predictor was used when fitting the models (see *Chapter 5* for more details on the principles of modelling non-linear relationships by using fractional polynomials in logistic regression analysis). The multivariable models were fitted in each of the 50 complete data sets and the estimated regression parameters (coefficients and variances) were combined using Rubin's rules.

After refitting the models, the same shrinkage method used in the development phase (see *Chapter 5* for details on the calculation of the heuristic shrinkage factor) was applied to correct the re-estimated intercepts and predictor coefficients (reduce model optimism). Finally, as with any newly developed prognostic model, updated models should also be externally validated. However, that was outside the scope of the SPRAINED study.

Results

Exploratory analysis

The study recruited a cohort of 682 participants across 10 EDs between 20 July 2015 and 17 March 2016. The flow of participants through the cohort study is detailed in *Figure 15*. Baseline characteristics of the SPRAINED observational cohort study participants are summarised in *Table 25*. On average, participants were slightly older in the SPRAINED cohort study than in CAST (33.62 years vs. 29.88 years, respectively). Participants in the SPRAINED cohort study had an average BMI in the overweight category (27.08 kg/m^2), similar to the CAST participants (26.34 kg/m^2). The mean pain scores when resting (38.5 points) or bearing weight on the ankle (71.3 points) of the SPRAINED cohort study participants were also very similar to those observed for the CAST participants (37.75 points when resting and 75.42 points when bearing weight). In contrast to CAST, in the SPRAINED cohort study about half of participants were female (52.05%), presented to an ED for assessment within 2 days of injury (90.03%) and were able to bear some weight on their injured ankles (73.56%).

Continuous predictor variables presented at least a minimal departure from a normal distribution, as evidenced in *Figures 16* and *17*. Some outliers were observed for participant age and BMI. However, all extreme values were clinically plausible, so no observations were dismissed. Correlations between predictors are presented in *Table 26*, ranging from very low values ($r = 0.011$ for BMI and ability to bear weight on the injured ankle) to moderate values ($r = 0.549$ for pain when resting and pain when bearing weight), which did not raise concerns about including them together in a multivariable model.

Events rates in the SPRAINED cohort study and CAST data sets for both outcomes, and the number of symptoms, at 9 months after injury, are described in *Table 27*. There was a lower rate of poor outcome for the SPRAINED cohort than for the CAST cohort.

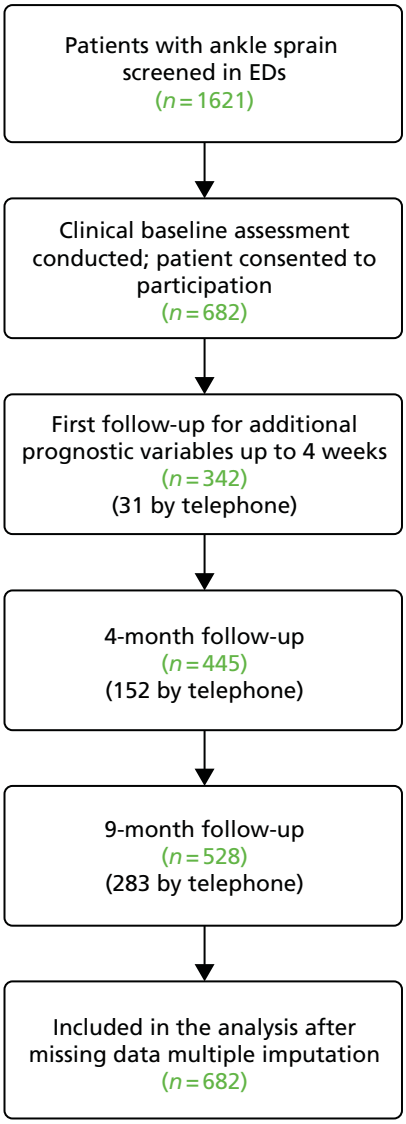


FIGURE 15 Flow of participants through the SPRAINED observational cohort study.

TABLE 25 Baseline characteristics of CAST and SPRAINED cohort study samples

	Trial/study			
	CAST		SPRAINED cohort	
Variable	Mean (SD)	Minimum, maximum	Mean (SD)	Minimum, maximum
Age (years)	29.88 (10.77)	16, 72	33.62 (13.38)	16, 89
Height (m)	1.73 (0.98)	1.47, 2.01	1.72 (1.02)	1.50, 2.01
Weight (kg)	78.56 (15.44)	39.92, 133.36	80.44 (18.13)	44.50, 180.00
BMI (kg/m^2)	26.34 (5.19)	16.07, 53.77	27.08 (5.70)	17.31, 64.30
Pain when resting (points)	37.75 (23.49)	0, 100	38.50 (22.50)	0, 100
Pain when bearing weight (points)	75.42 (19.61)	0, 100	71.30 (21.00)	0, 100

continued

TABLE 25 Baseline characteristics of CAST and SPRAINED cohort study samples (*continued*)

	Trial/study			
	CAST		SPRAINED cohort	
Variable	Mean (SD)	Minimum, maximum	Mean (SD)	Minimum, maximum
	Frequency	*%*	*Frequency*	*%*
Sex				
Male	337	57.71	327	47.95
Female	247	42.29	355	52.05
Days from injury to assessment				
0–2	118	44.87	614	90.03
≥ 3	145	55.13	68	9.97
Able to bear weight at baseline assessment				
No	446	77.03	179	26.44
Yes	133	22.97	498	73.56
Sprained the same ankle in the previous 12 months				
No	197	68.40	590	87.80
Yes	91	31.60	82	12.20
Sprained the same ankle at least twice before				
No	176	61.32	472	73.63
Yes	111	38.68	169	26.37
Recurrent sprain				
No	517	90.38	583	91.38
Yes	55	9.62	55	8.62
Current employment				
None	132	22.60	161	23.68
Part time	92	15.75	92	13.53
Full time	360	61.64	427	62.79
Injury mechanism				
At home	99	18.00	144	21.56
Practising sports	203	36.91	230	34.43
At work	79	14.36	91	13.62
Outside, in public	169	30.73	203	30.39
Ankle catching/locking				
Never	286	50.53	539	81.54
Rarely/sometimes	209	36.93	99	14.98
Often/always	71	12.54	23	3.48
Able to perform ankle ROM plantar flexion				
Always/often	101	17.84	170	25.30
Sometimes/rarely	247	43.64	230	34.23
Never	218	38.52	272	40.48

TABLE 25 Baseline characteristics of CAST and SPRAINED cohort study samples (*continued*)

	Trial/study			
	CAST		SPRAINED cohort	
Variable	Mean (SD)	Minimum, maximum	Mean (SD)	Minimum, maximum
Able to perform ankle ROM dorsiflexion				
Always/often	81	14.31	186	27.72
Sometimes/rarely	227	40.11	228	33.98
Never	258	45.58	257	38.30
Injury severity				
Grade 1	–	–	302	48.55
Grade 2	–	–	285	45.85
Grade 3	–	–	35	5.63

ROM, range of motion.

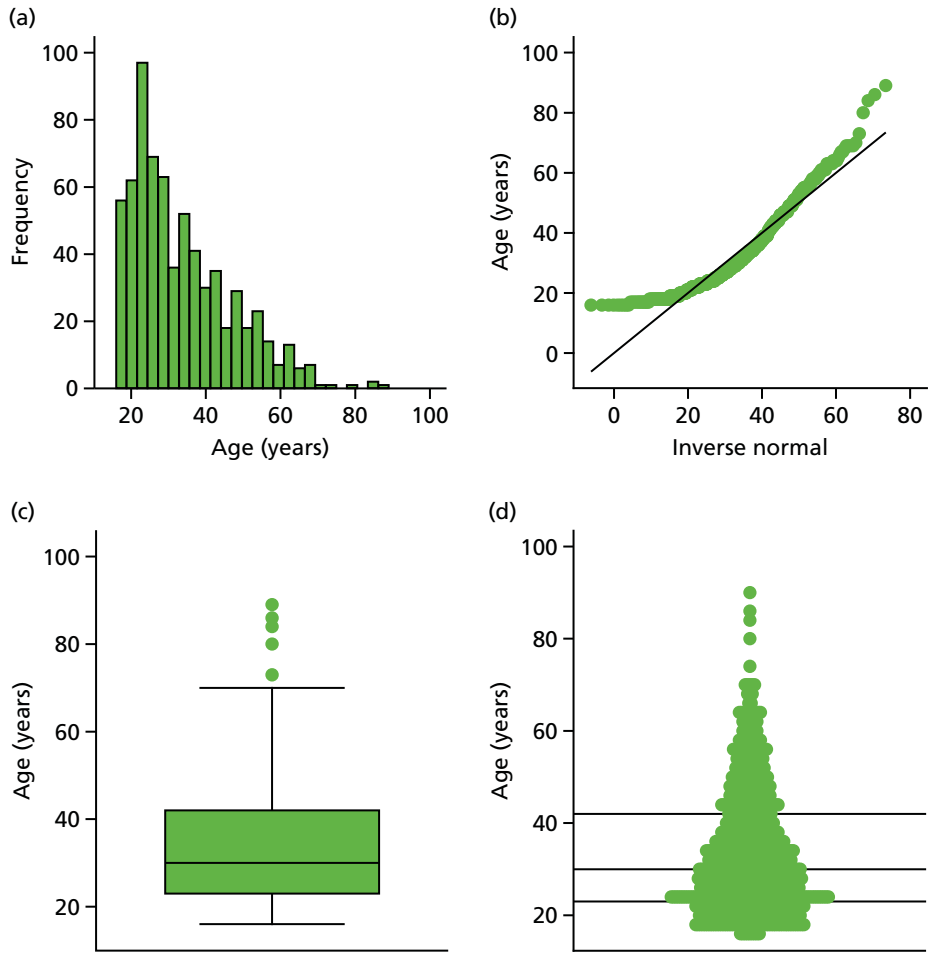


FIGURE 16 Distribution of age values in the SPRAINED cohort study data set. (a) Histogram; (b) normal plot; (c) box plot; and (d) dot plot.

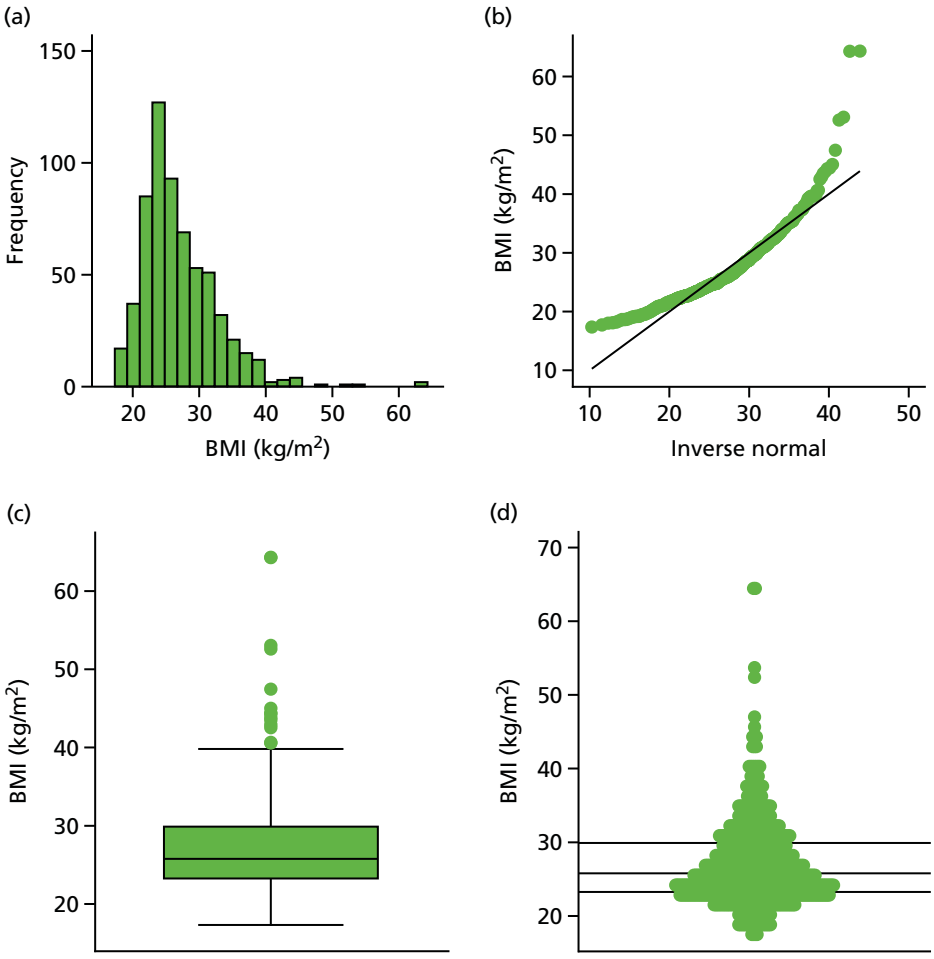


FIGURE 17 Distribution of BMI values in the SPRAINED cohort study data set. (a) Histogram; (b) normal plot; (c) box plot; and (d) dot plot.

TABLE 26 Spearman's rank-correlation coefficient matrix for all predictors included in the SPRAINED prognostic model for the risk of poor outcome 9 months after acute ankle injury

Variable	Age	BMI	Pain when		Days from injury to assessment	Able to bear weight on the injured ankle
			Resting	Bearing weight		
Age	–					
BMI	0.222	–				
Pain when resting	0.021	0.060	–			
Pain when bearing weight	–0.001	0.120	0.549	–		
Days from injury to assessment	0.083	0.047	–0.084	–0.116	–	
Able to bear weight on the injured ankle	0.050	0.011	–0.258	–0.393	0.110	–
Recurrent sprain	–0.127	–0.021	0.095	–0.031	0.053	–0.009

TABLE 27 Outcomes and component symptoms, *n* (%), in the CAST and SPRAINED data sets

	Symptoms/events								
Data set	Pain	Lack of confidence	General difficulty	Re-injury	Outcome 1	Missing	Outcome 2	Missing	Total,[a] *N*
CAST data set	84 (14.4)	42 (7.2)	67 (11.5)	46 (7.9)	116 (19.9)	144 (24.7)	140 (24.)	144 (24.7)	584
SPRAINED data set	3 (0.4)	23 (3.4)	37 (5.4)	78 (11.4)	46 (6.7)	155 (22.7)	109 (16.0)	150 (22.0)	682

a Total corresponds to number of participants recruited.

Model performance

The performance of the prediction models in the external validation data set (SPRAINED cohort study) was assessed in terms of calibration and discrimination. Calibration was graphically assessed with a calibration plot that showed calibration lines for each of the 50 imputed data sets, which was supplemented with the calibration slope and intercept. These parameters were first estimated with the original prognostic model, with poor outcome 9 months after ankle sprain (yes/no) as the outcome variable, and the linear predictor (log-odds) of the original prediction model (see *Chapter 5*, *Application of the SPRAINED study model* for the equation to calculate the linear predictor) as the only covariate.

Combined performance measures (by using Rubin's rules) are presented in *Table 28* and calibration plots overlaying the calibration lines from the 50 individual calibration plots are presented in *Figures 18* and *19*.

Overall, discrimination of the models for outcome 1 stayed fairly stable when compared with the performance of the model in the development data set: combined *c*-statistic 0.72 (95% CI 0.66 to 0.79). For outcome 2, a decrease in the discriminatory ability was noted: *c*-statistic 0.63 (95% CI 0.58 to 0.69).

TABLE 28 Summary of the combined performance measures (discrimination and calibration) for the prognostic models applied to the participants in the SPRAINED study sample

Model	*c*-statistic (95% CI)	Intercept (95% CI)	Slope (95% CI)
Outcome 1			
Baseline model	0.73 (0.66 to 0.79)	–0.91 (–1.18 to –0.65)	1.13 (0.76 to 1.50)
Updated model (baseline + 4-week predictors)	0.78 (0.72 to 0.84)	–0.62 (–0.89 to –0.34)	1.17 (0.86 to 1.48)
Baseline model applied to participants with moderate/severe injury (grades 2 and 3)	0.73 (0.64 to 0.81)	–1.13 (–1.53 to –0.73)	1.12 (0.55 to 1.69)
Updated model (baseline + 4-week predictors) applied to participants with moderate/severe injury (grades 2 and 3)	0.80 (0.72 to 0.88)	–0.85 (–1.25 to –0.44)	1.30 (0.81 to 1.78)
Outcome 2			
Baseline model	0.63 (0.58 to 0.69)	–0.25 (–0.44 to –0.06)	1.03 (0.65 to 1.42)
Updated model (baseline + 4-week predictors)	0.64 (0.59 to 0.69)	0.12 (–0.07 to –0.32)	0.68 (0.46 to 0.91)
Baseline model applied to participants with moderate/severe injury (grades 2 and 3)	0.62 (0.54 to 0.69)	–0.40 (–0.68 to –0.12)	0.94 (0.36 to 0.52)
Updated model (baseline + 4-week predictors) applied to participants with moderate/severe injury (grades 2 and 3)	0.63 (0.54 to 0.69)	–0.06 (–0.35 to 0.23)	0.65 (0.32 to 0.98)

Note
Performance measures for the different models are a combination of the individual estimates obtained from the analyses of the 20 imputed data sets. Estimates were combined by using Rubin's rules.

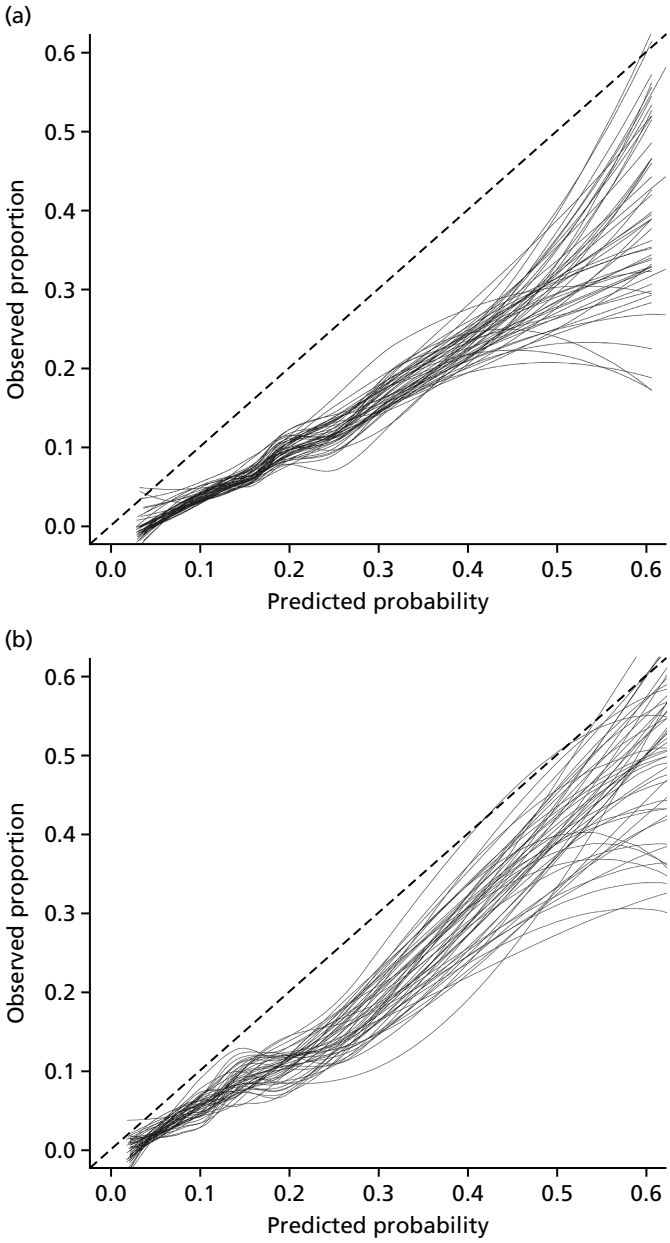


FIGURE 18 Calibration plots for the baseline and updated prognostic models to predict outcome 1, overlaying the 50 calibration lines derived from the individual imputed SPRAINED data sets. (a) Baseline; and (b) updated (baseline + 4-week predictors). Adapted with permission from Schlüssel *et al.*[26] © Author(s) (or their employer(s)) 2018. Re-use permitted under CC BY. Published by BMJ. This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: https://creativecommons.org/licenses/by/4.0/.

Calibration of the prognostic model in the external validation data set was poor for outcome 1, as can be evidenced by inspecting *Figure 19* (a calibration plot with overlaid calibration lines from the 50 imputed data sets). Well-calibrated models should produce calibration lines lying on (or at least close to) the 45° dashed line of perfect prediction (observed proportion and predicted probability matching perfectly). In this scenario, the calibration slope would be equal (or very close) to 1 and the calibration intercept equal (or very close) to 0. The combined calibration slope was > 1 (1.13, 95% CI 0.76 to 1.50) and the calibration intercept was smaller than zero (–0.91, 95% CI –1.18 to –0.65).

A calibration slope of > 1 indicates that the regression coefficients of the original model were too close to zero, which was the case after the correction for optimism (shrinkage) of the model. A calibration intercept

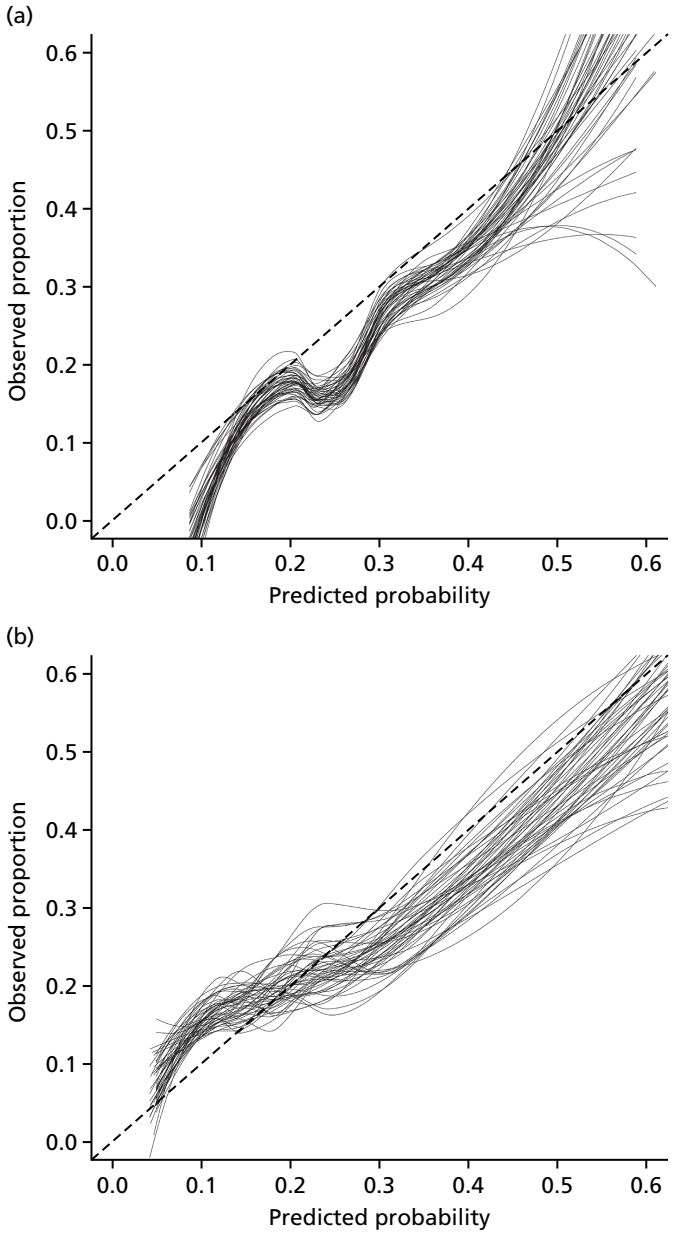


FIGURE 19 Calibration plots for the baseline and updated prognostic models to predict outcome 2, overlaying the 50 calibration lines derived from the individual imputed SPRAINED data sets. (a) Baseline; and (b) updated (baseline + 4-week predictors).

different from zero indicates that the model's predicted probabilities in the validation data set are systematically too high (intercept < 0) or too low (intercept > 0).

For the prognostic model developed to predict outcome 2, calibration was better than for the model to predict outcome 1 in terms of the calibration intercept (–0.25, 95% CI –0.44 to –0.06), and slope (1.03, 95% CI 0.65 to 1.42) (see *Table 28*). The updated model (baseline + 4-week predictors) for outcome 1 presented a better discriminatory ability in the SPRAINED data set than the baseline model (*c*-statistic = 0.78, 95% CI 0.72 to 0.84), but not better calibration in terms of intercept (–0.62, 95% CI –0.89 to –0.34). The same was observed for the updated model for outcome 2 (better discrimination but worse calibration) (see *Table 28*).

Table 29 shows how many of 1000 people would be identified as being at high risk of developing the outcome (based on thresholds of 5%, 10%, 15% and 20%), using the developed prognostic models, and how many of these would actually present poor outcome 9 months after an acute ankle sprain. There seems to be little difference between the baseline and updated models for outcome 1, with both models identifying

TABLE 29 Models performance (numbers at risk and outcomes identified) at varying risk thresholds for 1000 patients

	Outcome, *n*							
	1				2			
	Patient risk		Outcomes		Patient risk		Outcomes	
Predicted probability	High	Low	Identified	Not identified	High	Low	Identified	Not identified
Consider all high risk	1000	0	85	0	1000	0	198	0
Predicted probability as per baseline model								
≥ 5%	971	39	85	0	1000	0	198	0
≥ 10%	797	203	74	11	1000	0	198	0
≥ 15%	543	457	63	22	884	116	191	7
≥ 20%	351	649	52	33	636	364	138	60
Predicted probability as per updated model								
≥ 5%	882	118	85	0	993	7	198	0
≥ 10%	517	483	71	14	704	296	156	42
≥ 15%	358	642	56	29	456	544	106	92
≥ 20%	259	741	41	44	336	664	85	113

Note
Estimates based on complete-case analysis ($n = 271$ and $n = 283$ for outcomes 1 and 2, respectively).

a similar number of patients who experience a poor outcome after ankle sprain. However, fewer patients are deemed as being at high risk by using the updated model for outcome 1 (fewer false positives) across all thresholds of predicted probability, as estimated by the prognostic models. For outcome 2, the updated model misses more patients who actually develop the outcome (false negatives) when compared with the baseline model. Using either of the models seems to be beneficial when compared with not using any model (or considering all patients as being at high risk of developing poor outcome).

Subgroup analyses

As the prognostic models were developed using a data set from a clinical trial that included only participants with moderate or severe injuries (grades 2 or 3), it was decided that separate results on the models' performance would also be presented for a subgroup of participants classified according to their injury severity degree (grades 2 and 3).

Overall, both the calibration (intercepts and slopes) and discrimination (*c*-statistics) did not show any substantial improvement in the subgroup analysis for the baseline prognostic models to predict either outcome 1 or 2 (see *Table 28*). For the updated models (baseline + 4-week predictors), the intercept of the prognostic model to predict outcome 2 presented some improvement in terms of the calibration intercept, but not for the calibration slope (see *Table 28*).

Model recalibration

Before recalibrating the models, we considered investigating the predictive ability of two additional candidate predictors not included in the development phase (no data were available in the CAST data set), but for which information was collected at baseline in the SPRAINED cohort study: sprain severity and recovery

expectancy (time to recover from injury, as reported by the participants). Neither of the two variables showed statistically significant crude associations with the outcomes and presented very low predictive ability. For sprain severity, *c*-statistics were 0.48 (95% CI 0.40 to 0.57) and 0.50 (95% CI 0.44 to 0.56) for outcomes 1 and 2, respectively. For recovery expectancy, *c*-statistics were 0.56 (95% CI 0.48 to 0.64) and 0.50 (95% CI 0.44 to 0.55) for outcomes 1 and 2, respectively.

Results from the model update are presented in *Tables 30* and *31*. Predictor transformations were very similar to those observed for the original prognostic models developed with CAST data, apart from the fact that measures of pain were measured on a scale ranging from 0 to 10, and therefore an index was added, indicating that values derived from assessments conducted with the visual analogue scale (which ranges from 0 to 100) should be divided by 10 before any transformation is performed when applying the model to estimate individual risks (see *Table 30*). Coefficients obtained from the logistic regression models employed to update the models are presented in *Table 31*. Shrunk coefficients after applying the heuristic shrinkage factor to reduce optimism in the re-estimated model are also presented (see *Table 31*).

The results of the prognostic development (see *Chapter 5*) and validation are summarised and discussed together in *Chapter 7*.

TABLE 30 Transformations for non-categorical predictors in the recalibrated models for outcomes 1 and 2

	Outcome	
Variable	1	2
Age (years)	33.62	–
BMI (kg/m^2)	27.05	–
Pain when resting (score 0–100, divided by 10)	3.86	–3.86
Pain when bearing weight (score 0–100, divided by 10)	7.11	7.11

TABLE 31 Intercept and regression coefficients of the recalibrated prediction models for poor recovery 9 months after ankle sprain (outcomes 1 and 2), before and after correction for optimism (shrinkage)

	Outcome			
	1		2	
Predictor	Coefficient	Shrunk coefficient	Coefficient	Shrunk coefficient
Baseline model				
Age (years)	0.02	0.02	–	–
BMI (kg/m^2)	0.03	0.03	–	–
Pain when resting	0.19	0.17	0.07	0.06
Pain when bearing weight	0.18	0.16	0.10	0.09
> 2 days from injury to assessment	–0.88	–0.78	–0.62	–0.56
Able to bear weight on the injured ankle	–0.22	–0.19	–0.05	–0.04
Recurrent sprain	1.60	1.42	2.07	1.88
Intercept	–2.60	–2.52	–1.61	–1.57

continued

TABLE 31 Intercept and regression coefficients of the recalibrated prediction models for poor recovery 9 months after ankle sprain (outcomes 1 and 2), before and after correction for optimism (shrinkage) (*continued*)

	Outcome			
	1		2	
Predictor	Coefficient	Shrunk coefficient	Coefficient	Shrunk coefficient
Updated model (baseline + 4-week predictors)				
Age (years)	0.02	0.01	–	–
BMI (kg/m^2)	0.03	0.03	–	–
Pain when resting	0.17	0.15	0.06	0.05
Pain when bearing weight	0.14	0.12	0.08	0.07
Pain when bearing weight at 4 weeks after injury	0.03	0.03	0.01	0.01
> 2 days from injury to assessment	–1.23	–1.11	–0.71	–0.63
Able to bear weight on the injured ankle	–0.10	–0.09	0.07	0.06
Recurrent sprain	1.43	1.29	2.01	1.79
Intercept	–2.85	–2.73	–1.63	–1.58

Chapter 7 Overall discussion

The SPRAINED study research programme aimed to develop and externally validate prognostic models to aid clinical decision-making about the risk of poor outcome for people attending EDs with acute ankle sprains. The models were developed based on existing prognostic factor research (see *Chapter 3*) and expert consensus (see *Chapter 4*) and using a large cohort of multicentre RCT participants (see *Chapter 5*). The external validation of the model was assessed in a subsequent prospective observational cohort study (see *Chapter 6*). In this chapter, we consider the overall performance of the models, the limitations of the study and the implications for clinical practice and make recommendations for future research.

Performance of the SPRAINED prognostic models

Summary

The first prognostic model was developed to predict a composite outcome representing the presence of at least one of the following symptoms at 9 months after injury: persistent pain, functional difficulty or lack of confidence (outcome 1).

The second model was developed to predict a composite outcome representing the presence of at least one of the following symptoms or clinical events at 9 months after injury: persistent pain, functional difficulty, lack of confidence or recurrence of injury (outcome 2).

The models for outcome 1 and outcome 2 provided reasonable predictions of poor outcome for people with acute ankle sprain on the population used in their derivation (see *Chapter 5*).

There was a slight decrease in model discrimination for both models when evaluated in a prospectively collected external validation cohort study (see *Chapter 6*). The model for outcome 1 had better discrimination than the model for outcome 2. The variables for poor outcome used in model 1 (persistent pain, functional difficulty or lack of confidence) were, therefore, easier and more reliable to predict, and appear to have good clinical utility. Hence this would be the model of choice.

The model predicting presence of either persistent pain, functional difficulty, lack of confidence or recurrence of injury (outcome 2) showed good calibration, whereas there was miscalibration of the model predicting persistent pain, functional difficulty or lack of confidence (outcome 1).

Updating these models, which used baseline data collected at the ED, with an additional variable at 4 weeks after the injury (pain when bearing weight on the ankle) improved the discriminatory ability and apparent calibration. However, improvements in model performance were modest. Balancing the practical challenges and resource implications of obtaining additional data at 4 weeks after presentation at the ED with the improvements in prediction is likely to be an important consideration when selecting a model for use in clinical practice.

Despite some miscalibration of the models, the external validation study (see *Chapter 6*) found that the model performance was reasonable for identifying patients at increased risk of poor outcome after acute ankle sprain, and showed benefit when compared with not using any model. To the best of our knowledge, there are no other prognostic models that have been developed and externally validated using robust methods for this patient group (see *Chapter 3*). The SPRAINED prognostic models may assist clinical decision-making when assessing and advising people with ankle sprains in the ED setting and when deciding on ongoing management. The models benefit from using predictors that are simple to obtain during routine clinical assessment. Recalibration of the models may be required to improve the accuracy of the predicted risks in other populations (both in and outside the UK).

Differences in prognostic model performance in the development and external validation studies

The differences in model performance between the development and external validation studies could have several explanations. First, any prognostic model is expected to perform better in the data set used in its development. Second, the very nature of the two studies can explain, in part, the poor calibration of the model, as the development data set derived from a RCT, whereas the external validation data set was from a prospective observational cohort study with less restrictive eligibility criteria. The aim of the observational cohort study was to be representative of the general population seeking medical assistance for acute ankle sprains at EDs in the UK NHS. Third, the case mix in the two data sets might also explain the differences in model performance, as some of the most important predictors (e.g. number of days from injury to assessment and ability to bear weight on the injured ankle) were not equally distributed among participants in the two data sets. Finally, the differences in the outcomes' rates (particularly for outcome 1) might have influenced the poor calibration of the models observed for the SPRAINED cohort study. We recommend that the recalibrated prognostic models should be evaluated in different sets of patients.

An exhaustive set of predictors was used, which included clinical consensus to gain insight into what factors are easy to implement and acceptable. Physical tests could not be included as there were insufficient data, although these have not appeared to be useful tests in previous evaluations. It might be that in the future new data, such as MRI or simple gait analysis, will be able to add extra prognostic information. Education was excluded from our considerations, but, given its low priority and relatively low contribution to only one model, it is unlikely to provide much additional prognostic information.

The consensus group (see *Chapter 4*) suggested that psychological variables may improve the prediction, and, although an additional variable was collected on the participant's expectation about recovery, there was limited evidence that this additional variable had prognostic utility.

Strengths and limitations

To the best of our knowledge, this is the first study to (1) develop a prognostic model to predict poor outcome in people with acute ankle sprains using an adequately large cohort to explore a wide range of clinically plausible candidate predictors, (2) use robust statistical methods to assess the performance of the prognostic models and (3) include a large prospective cohort study to enable external validation. We needed to conduct the observational cohort as there were no other available and sufficiently large data sets with data on a wide range of candidate predictors available for an external validation. Generalisability of the findings are enhanced by the multicentre data from the CAST and SPRAINED cohort studies that represented a range of district general and major trauma centres.

We followed the most recent guidelines available on the reporting of prognostic model development and used methods that, to the best of our knowledge, are the most widely recommended. For example, continuous variables, whenever possible, were kept as continuous, to avoid loss of information. Non-linear relationships were investigated using the best variable transformations found by MFPs. The study included an internal correction for model optimism (shrinkage of regression coefficients and intercepts), as well as an external validation phase. Missing data are almost inevitable in studies of this nature; however, the number of missing data in the external validation data set was considerably smaller than that observed in the development data set, and missing data imputation was also used to produce a set of 50 complete data sets, which enabled more robust analyses.

The SPRAINED study has limitations that must be considered when interpreting the results described in this report. First, the data used to develop the two proposed prognostic models were from a prior RCT (CAST), so were not originally intended to fulfil this aim. However, the CAST cohort did represent the best data set available, with data on the symptoms and clinical events of interest to compose the two outcomes for the SPRAINED prognostic model, and for the majority of the candidate prognostic variables considered to have

predictive ability at the time of the study's conception. CAST was a pragmatic RCT, with relatively open eligibility criteria, that aimed to investigate the effect of four different interventions on a different set of (primary and secondary) outcomes. The CAST data set was not optimally sized for developing prognostic models; had it been larger, it might have provided more robust estimates, resulting in models with less optimism. As previously highlighted, the low EPV observed for the two models developed might have contributed to the optimism found for both prognostic models and, therefore, to the poor calibration on the external validation data set. Finally, another important limitation relates to the number of missing data observed in the development data set. Because of the number of missing data, some of the candidate predictors had to be dropped before the process of data imputation because the number of missing observations (> 60%) was considered too high. Therefore, some important predictors could conceivably have been missed in the development phase of the SPRAINED study.

A key focus of the SPRAINED study was that the prognostic factor variables needed to be based on routinely collected clinical information. It is possible that information from imaging techniques, such as MRI, could have resulted in a more accurate estimation of risk (see *Chapter 3*). However, this type of investigation is not routinely used or available in the context of an ED consultation. We therefore limited our investigation to prognostic factors that are or could easily be obtained during a routine assessment of a person with an acute ankle sprain in the ED.

The rates of poor outcome in the SPRAINED cohort study were lower than in CAST (7% vs. 20% for outcome 1 and 24% vs. 16% for outcome 2) and lower than the rates of approximately 30% reported in previous systematic reviews.[3,4] These variations in poor outcome rates highlight the potential issue of different sampling frames. It could be argued that the observational cohort which we recruited for SPRAINED was a reasonable representation of the rates of poor outcome in patients presenting to EDs in the UK, as all types of adult patients with an ankle sprain were included, there was low participant burden from participation compared with many clinical trials and we achieved good levels of follow-up.

Other prognostic models reported during the SPRAINED study

Our systematic review of the literature highlighted limitations in the evidence relating to predictive factors for recovery from ankle sprain. Since this review, Doherty *et al.*[79] have reported on movement tests performed at 2 weeks after injury as predictors of CAI after acute ankle sprain. They found that inability to complete two out of five dynamic movement tests had a sensitivity of 83% and specificity of 55% for identifying those classified as having CAI.[79] These assessments are not currently routinely available clinical information in most EDs in the UK; however, these results may indicate that consideration of predictive factors in later stages of recovery may be appropriate.

Clinical implications of the SPRAINED study

Estimating the risk of a poor outcome for a person attending an ED with an ankle sprain is desirable because of the large number of individuals presenting with these injuries and the difficulty in determining who will struggle to recover. Many people present in the acute phase with a degree of ankle pain, swelling, loss of motion and difficulty bearing weight on the injured leg. Clinical examination is often challenging, as tolerance of physical examination tests is limited by pain and the examinations have been found to have poorer sensitivity and specificity within the first 48 hours after injury than 5 days after injury.[80] As a result, it is difficult to decide who may benefit from monitoring or rehabilitation. The value of a prognostic model is evident, but in order for it to be utilised in clinical practice, it needs to be quick and simple to use, and offer a sufficiently accurate estimation of risk of poor outcome to be clinically worthwhile.

The prognostic models have the potential to assist clinicians to decide whether or not an early review is merited and to offer some reassurance that people who are not followed up are likely to be on a positive recovery trajectory. As with other prognostic models, any potential benefits from being able to estimate an outcome should be considered in the context of the performance of the models and the potential risks of an inaccurate prediction for the person being assessed. Given some limitations in predictive performance of the SPRAINED prognostic models at the development (see *Chapter 5*) and external validation (see *Chapter 6*) stages, we suggest that their value would be in assisting the clinician in estimating the probability of a poor outcome, rather than being a decision-making tool in isolation. If implemented in clinical practice, it should be noted that there is a degree of uncertainty in the calculated risk of poor outcome when using the SPRAINED prognostic models. This uncertainty in estimation could lead to over- or under-referral of patients to review clinics or treatment, such as physiotherapy, and highlights the caution required in using the calculated individual risks when counselling patients about their prognosis. Further research is recommended to evaluate the impact of using the SPRAINED prognostic models on clinical practice and patient outcomes, and to assess the acceptability and uptake of use by ED clinicians.

Of note, 78 out of 682 (11%) participants reported a recurrence of sprain within 9 months of their initial presentation in the external validation study. It could be argued that widening the classification to recurrence of sprain is more consistent with existing definitions of CAI.[81] Although we did not set out to predict CAI specifically, we recognise that people with a poor outcome, as defined by the SPRAINED study, would probably include patients with this condition.

One of the important aspects of assessing the clinical usefulness of a multivariable prognostic model is that it is a better predictor of poor outcome than the overall clinical impression of clinical severity of the presenting ankle sprain. Future work could examine how well the model performs in comparison to the clinician impression.

Implementation of the SPRAINED prognostic models

Other prediction models are in routine clinical use in the ED. One prediction model being used routinely is for ankle injuries, the Ottawa ankle rules;[82] these are used to help determine which patients should be considered for radiographs to rule out a fracture.[83] Patients entered into the SPRAINED study would have been assessed to rule out a fracture during their ED assessment. We envisage that implementation of the SPRAINED prognostic model could also be used in the assessment of this patient group, once the clinician is satisfied that there is no fracture.

An application of the SPRAINED prognostic models that we recommend for future investigation is whether or not the models can be used to stratify patients to post-injury interventions that are matched to the level of risk of poor outcome. There have been inconsistencies in the findings of trials investigating the effectiveness of physiotherapy rehabilitation after acute ankle sprain.[84,85] We hypothesise that, as most patients attending the ED have a good prognosis, better targeting of higher-intensity interventions to those at greater risk of poor outcome may enhance the clinical effectiveness and cost-effectiveness of rehabilitation; however, this requires formal evaluation.

The prognostic model requires a calculation too complex for easy use in the clinical setting, so it would require a computer application to facilitate the calculation of probability for poor outcome for the person being examined in the ED. A web-based calculator or application could be developed specifically for the SPRAINED prognostic models; this is an area of work that will be taken forward by the SPRAINED investigators. Owing to limitations in the performance of the models, an issue to address when presenting the calculated risks to clinicians will be to concurrently make users aware of the prediction accuracy.

Recommendations for future research

Further research is recommended to:

- determine appropriate cut-off points or score ranges from the prognostic model for identifying patients more likely to benefit from different clinical pathways
- assess whether or not the prognostic model can improve decision-making and targeting of treatment, and ultimately patient outcomes
- evaluate the acceptability and uptake of use by ED clinicians
- examine how well the model performs in comparison with clinician impression on prognosis and assessment of clinical severity of the presenting ankle sprain
- investigate whether or not a wider range of psychological, or other types of variables that were not included in the SPRAINED study, improve prediction.

It was also noted that recalibration of the models may be required to improve the accuracy of the predicted risks in other populations (in and outside the UK).

Conclusions

The SPRAINED study research programme aimed to develop and externally validate prognostic models to aid clinical decision-making about the risk of poor outcome for people attending EDs with an acute ankle sprain. The models were developed based on existing prognostic factor research and expert consensus and using a large cohort of multicentre RCT participants. The external validation of the model was assessed in a subsequent prospective observational cohort study.

The SPRAINED prognostic models performed reasonably and showed benefit in identifying patients who are at a high risk of poor outcome after acute ankle sprain when compared with not using any model (consider all patients as being at high risk of poor outcome), so may assist clinical decision-making when assessing and advising people with ankle sprains in the ED setting and when deciding on ongoing management. The models benefit from using predictors that are simple to obtain during routine clinical assessment.

Further research to evaluate the performance of the models in other settings is recommended. Further refinement of the models, including external validation of the recalibrated models or identifying additional predictors, may be required. The impact of implementing and using either model in clinical practice, in terms of acceptability and uptake by ED staff and their impact on patient outcomes, should also be investigated.

Acknowledgements

SPRAINED study team

- Chief investigator: Sarah E Lamb.
- Study lead: David J Keene.
- Co-investigators: Gary S Collins, Mark A Williams, Steve Goodacre, Matthew Cooke, Stephen Gwilym, Philip Hormbrey, David Wilson, Jennifer Bostock.
- Study co-ordinator and administrator: Daryl A Hagan.
- Senior study manager: Damian Haywood.
- Research Physiotherapists: Jacqueline Thompson, Christopher Byrne.
- Study statisticians: Michael M Schlüssel, Gary S Collins.

TABLE 32 Principal investigators/research fellows, nurses or therapists by hospital site, in order of date starting recruitment

Principal investigator	Research nurses, therapists and associates	Hospital name	NHS trust name
Dr Philip Hormbrey	Sally Beer, Amanda Budden, Alexis Espinosa, Dominique Georgiou, Louise Findlay	John Radcliffe Hospital	Oxford University Hospital NHS Foundation Trust
Dr Susan Dorrian	Samantha Stafford, Nathan Humphries, David Hunt	Heartlands Hospital and Solihull Hospital	Heart of England NHS Foundation Trust
Professor Steve Goodacre	Rachel Walker, Anna Wilson, Nicola Hindmarch, Craig Jones, Zoe Dutton, John Parry, Charlotte Green	Northern General Hospital	Sheffield Teaching Hospitals NHS Foundation Trust
Dr Victoria Stacey	Claire Hunt, Natalie Bynorth, Pauline Brown, Kayleigh Collins, Estelle Nambela	Cheltenham General Hospital and Gloucester Royal Hospital	Gloucestershire Hospital NHS Foundation Trust
Professor Tim Coats	Lisa McClelland, Elisabeth Cadman-Moore	Leicester Royal Infirmary	University Hospitals of Leicester NHS Trust
Dr Sarah Wilson	Louise Chandler, Louise Foster, Vikki Diduca, Joana Da Rocha	Wexham Park Hospital	Frimley Health NHS Foundation Trust
Dr Jason Kendall	Lee Cameron, Rachel Ozanne, Sue Kempson, Ruth Worner, Beverley Faulkner, Caroline Ellis	Southmead Hospital	North Bristol NHS Trust
Dr David Clarke	Nicola Jacques, Dariusz Pabianczyk, Ria Diel, Andrzej Adamowicz, Abby Brown, Claire Burnett, Daniel Sedgewick, Claire Sayner, Jane Macpherson, Elizabeth Oastler, Mitzi Baylis, Caroline Lewis, Helen Ingolfsrid, Rikki Davies, Carys Davies, Teresa Hobbs	Royal Berkshire Hospital	Royal Berkshire NHS Foundation Trust
Ms Antoanela Colda	Gill Ritchie, Seema Chavda	Milton Keynes University Hospital	Milton Keynes University Hospital NHS Foundation Trust
Dr Deborah Mayne	Jackie Berry, Sarah Patch, Julie Camsooksai, Lee Tbaily	Poole Hospital	Poole Hospital NHS Foundation Trust

Study Steering Committee

Professor Richard Riley (chairperson), Professor Kevin Mackway-Jones and Professor Suzanne McDonough.

Other acknowledgements

Special thanks to the Centre for Health, Law and Emerging Technologies (HeLEX) for their collaboration on the Dynamic Consent pilot study, in particular Professor Jane Kaye, Harriet Teare and Jeremy Holland.

We also recognise the contributions of the following individuals at OCTRU and the Centre for Rehabilitation Research for their support in delivering the SPRAINED study: Vicki Barber, Lesley Morgan, Emma Roberts, Scott Parsons, Sue Davolls, Katie Chegwin, Emma Haines, Oliver Conway, Hannah Ashby, Asima Qayyum, Tim Cranston, Patrick Julier, Lucy Eldridge, Simon Shayler, Joanna Black, Deborah Brown and others who have provided advice and support throughout the course of the study.

Participants in the SPRAINED study consensus meeting

Chrissy Aimes (Paramedic, South Central Ambulance Service), Emma Batchelor (Physiotherapist, University Hospital Birmingham NHS Foundation Trust), Emma Bolton (ED Advanced Nurse Practitioner, Gloucestershire Hospitals NHS Foundation Trust), Mrs Jennifer Bostock (PPI lead/co-applicant), Ms Lucy Cameron (Paramedic, South Central Ambulance Service), Dr David Clarke (ED Consultant, Royal Berkshire NHS Trust), Professor Matthew Cooke (ED Consultant, Heart of England Foundation Trust), Mr Jason Franks (Paramedic, South Central Ambulance Service), Professor Steve Goodacre (ED Consultant, Sheffield Teaching Hospitals NHS Foundation Trust), Dr Philip Hormbrey (ED Consultant, Oxford University Hospital NHS Foundation Trust), Mr Nathan Humphries [Extended Scope Practitioner (ESP) Physiotherapist, Heart of England Foundation Trust], Mr David Hunt (ESP Physiotherapist, Heart of England Foundation Trust), Mrs Claire Hunt (ESP Physiotherapist, Gloucestershire Hospitals NHS Foundation Trust), Dr Liza Keating (ED Consultant, Royal Berkshire NHS Trust), Dr David Keene (Clinical Researcher, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford), Mrs Sue Kempson (ESP Physiotherapist, North Bristol NHS Trust), Professor Sallie Lamb (Clinical Researcher, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford) and Dr David Wilson (Consultant Radiologist, St Luke's Radiology).

Funding

The SPRAINED study was funded by the National Institute for Health Research (NIHR) Health Technology Assessment programme (project number 13/19/06); and supported by the NIHR Biomedical Research Centre, Oxford, and the NIHR Fellowship programme (Dr David J Keene, PDF-2016-09-056). Sarah E Lamb receives funding from the NIHR Collaboration for Leadership in Applied Health Research and Care Oxford at Oxford Health NHS Foundation Trust.

Contributions of authors

David J Keene (Research Fellow in Trauma Rehabilitation/NIHR Postdoctoral Research Fellow) was the study lead, led the development and authorship of the report and was responsible for the overall management of the project.

Michael M Schlüssel (Medical Statistician) developed and carried out data analysis, co-authored the report and provided statistical input throughout the study.

Jacqueline Thompson (Research Physiotherapist) led the systematic review, provided training and clinical support to collaborating sites, facilitated follow-up of participants, and co-produced the consensus meeting chapter.

Daryl A Hagan (Study Co-ordinator and Administrator) was responsible for the day-to-day co-ordination of the project, data collection, queries and data cleaning, and collation and editing of report chapters.

Mark A Williams (Senior Lecturer in Physiotherapy and Rehabilitation) was the previous study lead, led the consensus meeting process and co-produced the consensus meeting chapter.

Christopher Byrne (Lecturer in Physiotherapy) produced the systematic review, and provided training and clinical support to collaborating sites.

Steve Goodacre (Professor of Emergency Medicine) provided academic expertise and advice at key points, was a recruiting site principal investigator and reviewed the report.

Matthew Cooke (Professor of Emergency Medicine) provided academic expertise and advice at key points and reviewed the report.

Stephen Gwilym (Consultant Surgeon and Honorary Senior Lecturer) provided academic expertise and advice at key points and reviewed the report.

Philip Hormbrey (Consultant in Emergency Medicine) provided academic expertise and advice at key points, was a recruiting site principal investigator and reviewed the report.

Jennifer Bostock (PPI Representative) provided consultation and key input of patient and public perspective throughout study and reviewed the report.

Kirstie Haywood (Senior Research Fellow in Patient Reported Outcomes) provided consultation and senior facilitation of the consensus meeting process.

David Wilson (Honorary Consultant Radiologist) provided clinical expertise and advice at key points and reviewed the report.

Gary S Collins (Professor of Medical Statistics) was responsible for the study design, supervised data analysis, provided academic expertise and advice throughout project and co-authored the report.

Sarah E Lamb (Professor of Trauma Rehabilitation/Director of Oxford Clinical Trials Research Unit) was the chief investigator and had overall responsibility for the study, the design, academic leadership and authorship of the report.

Publications

Thompson JY, Byrne C, Williams MA, Keene DJ, Schlüssel MM, Lamb SE. Prognostic factors for outcome following acute lateral ankle ligament sprain. A systematic review. *BMC Musculoskelet Disord* 2017;**18**:421.

Schlüssel MM, Keene DJ, Collins GS, Bostock J, Byrne C, Goodacre S, *et al.* Development and prospective external validation of a tool to predict poor recovery at 9 months after acute ankle sprain in UK emergency departments: the SPRAINED prognostic model. *BMJ Open* 2018;**8**:e022802.

Data-sharing statement

All data requests should be submitted to the corresponding author for consideration. Access to anonymised data may be granted following review. Exclusive use will be retained until the publication of major outputs.

Patient data

This work uses data provided by patients and collected by the NHS as part of their care and support. Using patient data is vital to improve health and care for everyone. There is huge potential to make better use of information from people's patient records, to understand more about disease, develop new treatments, monitor safety, and plan NHS services. Patient data should be kept safe and secure, to protect everyone's privacy, and it's important that there are safeguards to make sure that it is stored and used responsibly. Everyone should be able to find out about how patient data are used. #datasaveslives You can find out more about the background to this citation here: https://understandingpatientdata.org.uk/data-citation.

References

1. Cooke MW, Lamb SE, Marsh J, Dale J. A survey of current consultant practice of treatment of severe ankle sprains in emergency departments in the United Kingdom. *Emerg Med J* 2003;**20**:505–7. https://doi.org/10.1136/emj.20.6.505
2. Martin RL, Davenport TE, Paulseth S, Wukich DK, Godges JJ, Orthopaedic Section American Physical Therapy Association. Ankle stability and movement coordination impairments: ankle ligament sprains. *J Orthop Sports Phys Ther* 2013;**43**:A1–40. https://doi.org/10.2519/jospt.2013.0305
3. van Rijn RM, van Os AG, Bernsen RM, Luijsterburg PA, Koes BW, Bierma-Zeinstra SM. What is the clinical course of acute ankle sprains? A systematic literature review. *Am J Med* 2008;**121**:324–31.e6. https://doi.org/10.1016/j.amjmed.2007.11.018
4. Polzer H, Kanz KG, Prall WC, Haasters F, Ockert B, Mutschler W, Grote S. Diagnosis and treatment of acute ankle injuries: development of an evidence-based algorithm. *Orthop Rev* 2012;**4**:e5. https://doi.org/10.4081/or.2012.e5
5. Cooke MW, Marsh JL, Clark M, Nakash R, Jarvis RM, Hutton JL, *et al.* Treatment of severe ankle sprain: a pragmatic randomised controlled trial comparing the clinical effectiveness and cost-effectiveness of three types of mechanical ankle support with tubular bandage. The CAST trial. *Health Technol Assess* 2009;**13**(13). https://doi.org/10.3310/hta13130
6. Verhagen RA, de Keizer G, van Dijk CN. Long-term follow-up of inversion trauma of the ankle. *Arch Orthop Trauma Surg* 1995;**114**:92–6. https://doi.org/10.1007/BF00422833
7. Wikstrom EA, Hubbard-Turner T, McKeon PO. Understanding and treating lateral ankle sprains and their consequences: a constraints-based approach. *Sports Med* 2013;**43**:385–93. https://doi.org/10.1007/s40279-013-0043-z
8. Stiell I, Wells G, Laupacis A, Brison R, Verbeek R, Vandemheen K, Naylor CD. Multicentre trial to introduce the Ottawa ankle rules for use of radiography in acute ankle injuries. Multicentre Ankle Rule Study Group. *BMJ* 1995;**311**:594–7. https://doi.org/10.1136/bmj.311.7005.594
9. van Dijk CN, Mol BW, Lim LS, Marti RK, Bossuyt PM. Diagnosis of ligament rupture of the ankle joint. Physical examination, arthrography, stress radiography and sonography compared in 160 patients after inversion trauma. *Acta Orthop Scand* 1996;**67**:566–70. https://doi.org/10.3109/17453679608997757
10. Hemingway H, Croft P, Perel P, Hayden JA, Abrams K, Timmis A, *et al.* Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ* 2013;**346**:e5595. https://doi.org/10.1136/bmj.e5595
11. Hingorani AD, Windt DA, Riley RD, Abrams K, Moons KG, Steyerberg EW, *et al.* Prognosis research strategy (PROGRESS) 4: stratified medicine research. *BMJ* 2013;**346**:e5793. https://doi.org/10.1136/bmj.e5793
12. Riley RD, Hayden JA, Steyerberg EW, Moons KG, Abrams K, Kyzas PA, *et al.* Prognosis Research Strategy (PROGRESS) 2: prognostic factor research. *PLOS Med* 2013;**10**:e1001380. https://doi.org/10.1371/journal.pmed.1001380
13. Hiller CE, Nightingale EJ, Lin CW, Coughlan GF, Caulfield B, Delahunt E. Characteristics of people with recurrent ankle sprains: a systematic review with meta-analysis. *Br J Sports Med* 2011;**45**:660–72. https://doi.org/10.1136/bjsm.2010.077404
14. Linde F, Hvass I, Jürgensen U, Madsen F. Early mobilizing treatment in lateral ankle sprains. Course and risk factors for chronic painful or function-limiting ankle. *Scand J Rehabil Med* 1986;**18**:17–21.

15. Akacha M, Hutton J, Lamb SE. *Modelling Treatment, Age- and Gender-specific Recovery in Acute Injury Studies*. Working Paper No. 9. Coventry: Centre for Research in Statistical Methodology, University of Warwick; 2010.
16. Kerkhoffs GM, Rowe BH, Assendelft WJ, Kelly KD, Struijs PA, van Dijk CN. WITHDRAWN: Immobilisation and functional treatment for acute lateral ankle ligament injuries in adults. *Cochrane Database Syst Rev* 2013;**3**:CD003762. https://doi.org/10.1002/14651858.CD003762.pub2
17. Kerkhoffs GM, Handoll HH, de Bie R, Rowe BH, Struijs PA. Surgical versus conservative treatment for acute injuries of the lateral ligament complex of the ankle in adults. *Cochrane Database Syst Rev* 2007;**2**:CD000380. https://doi.org/10.1002/14651858.CD000380.pub2
18. Lamb SE, Marsh JL, Hutton JL, Nakash R, Cooke MW, Collaborative Ankle Support Trial (CAST Group). Mechanical supports for acute, severe ankle sprain: a pragmatic, multicentre, randomised controlled trial. *Lancet* 2009;**373**:575–81. https://doi.org/10.1016/S0140-6736(09)60206-3
19. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot Ankle Int* 2001;**22**:788–94. https://doi.org/10.1177/107110070102201004
20. Brooks R. EuroQol: the current state of play. *Health Policy* 1996;**37**:53–72. https://doi.org/10.1016/0168-8510(96)00822-6
21. People in Research. *Opportunities for Public Involvement in NHS, Public Health and Social Care Research*. URL: www.peopleinresearch.org/ (accessed August 2017).
22. World Medical Association. *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*. URL: www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/ (accessed August 2017).
23. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). *Efficacy Guidelines*. 2016. URL: www.ich.org/products/guidelines/efficacy/article/efficacy-guidelines.html (accessed August 2017).
24. NHS Health Research Authority. *UK Framework for Health and Social Care Research*. London: Health Research Authority; 2018. URL: www.hra.nhs.uk/planning-and-improving-research/policies-standards-legislation/uk-policy-framework-health-social-care-research/ (accessed August 2017).
25. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, *et al.* Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;**162**:W1–73. https://doi.org/10.7326/M14-0698
26. Schlüssel MM, Keene DJ, Collins GS, Bostock J, Byrne C, Goodacre S, *et al.* Development and prospective external validation of a tool to predict poor recovery at 9 months after acute ankle sprain in UK emergency departments: the SPRAINED prognostic model. *BMJ Open* 2018;**8**:e022802. https://doi.org/10.1136/bmjopen-2018-022802
27. Williams M, Thompson J, Collins G, Schlussel M, Lamb S. *Prognostic Factors for Outcome Following Acute Ankle Ligament Sprain: A Systematic Review*. PROSPERO; 2014. CRD42014014471. URL: www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42014014471 (accessed August 2017).
28. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan – a web and mobile app for systematic reviews. *Syst Rev* 2016;**5**:210. https://doi.org/10.1186/s13643-016-0384-4
29. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. *Ann Intern Med* 2013;**158**:280–6. https://doi.org/10.7326/0003-4819-158-4-201302190-00009

30. de Bie RA, de Vet HC, van den Wildenberg FA, Lenssen T, Knipschild PG. The prognosis of ankle sprains. *Int J Sports Med* 1997;**18**:285–9. https://doi.org/10.1055/s-2007-972635

31. Wilson RW, Gansneder BM. Measures of functional limitation as predictors of disablement in athletes with acute ankle sprains. *J Orthop Sports Phys Ther* 2000;**30**:528–35. https://doi.org/10.2519/jospt.2000.30.9.528

32. Cross KM, Worrell TW, Leslie JE, Van Veld KR. The relationship between self-reported and clinical measures and the number of days to return to sport following acute lateral ankle sprains. *J Orthop Sports Phys Ther* 2002;**32**:16–23. https://doi.org/10.2519/jospt.2002.32.1.16

33. Langner I, Frank M, Kuehn JP, Hinz P, Ekkernkamp A, Hosten N, Langner S. Acute inversion injury of the ankle without radiological abnormalities: assessment with high-field MR imaging and correlation of findings with clinical outcome. *Skeletal Radiol* 2011;**40**:423–30. https://doi.org/10.1007/s00256-010-1017-y

34. van Middelkoop M, van Rijn RM, Verhaar JA, Koes BW, Bierma-Zeinstra SM. Re-sprains during the first 3 months after initial ankle sprain are related to incomplete recovery: an observational study. *J Physiother* 2012;**58**:181–8. https://doi.org/10.1016/S1836-9553(12)70109-1

35. van der Wees P, Hendriks E, van Beers H, van Rijn R, Dekker J, de Bie R. Validity and responsiveness of the ankle function score after acute ankle injury. *Scand J Med Sci Sports* 2012;**22**:170–4. https://doi.org/10.1111/j.1600-0838.2010.01243.x

36. O'Connor SR, Bleakley CM, Tully MA, McDonough SM. Predicting functional recovery after acute ankle sprain. *PLOS ONE* 2013;**8**:e72124. https://doi.org/10.1371/journal.pone.0072124

37. Medina McKeon JM, Bush HM, Reed A, Whittington A, Uhl TL, McKeon PO. Return-to-play probabilities following new versus recurrent ankle sprains in high school athletes. *J Sci Med Sport* 2014;**17**:23–8. https://doi.org/10.1016/j.jsams.2013.04.006

38. Olerud C, Molander H. A scoring scale for symptom evaluation after ankle fracture. *Arch Orthop Trauma Surg* 1984;**103**:190–194.

39. van Ochten JM, Mos MCE, van Putte-Katier N, Oei EHG, Bindels PJE, Bierma-Zeinstra SMA, *et al.* Structural abnormalities and persistent complaints after an ankle sprain are not associated: an observational case control study in primary care. *Br J Gen Pract* 2014;**64**:e545–53. https://doi.org/10.3399/bjgp14X681349

40. Royston P, Moons KG, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *BMJ* 2009;**338**:b604. https://doi.org/10.1136/bmj.b604

41. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *J Clin Epidemiol* 2015;**68**:112–21. https://doi.org/10.1016/j.jclinepi.2014.11.010

42. Black N, Murphy M, Lamping D, McKee M, Sanderson C, Askham J, Marteau T. Consensus development methods: a review of best practice in creating clinical guidelines. *J Health Serv Res Policy* 1999;**4**:236–48. https://doi.org/10.1177/135581969900400410

43. Delbecq AL, Van de Ven AH, Gustafson DH. *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*. Glenview, IL: Scott Foresman Company; 1975.

44. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, *et al.* Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol* 2014;**67**:745–53. https://doi.org/10.1016/j.jclinepi.2013.11.013

45. Harvey N, Holmes CA. Nominal group technique: an effective method for obtaining group consensus. *Int J Nurs Pract* 2012;**18**:188–94. https://doi.org/10.1111/j.1440-172X.2012.02017.x

46. Haywood KL, Griffin XL, Achten J, Costa ML. Developing a core outcome set for hip fracture trials. *Bone Joint J* 2014;**96–B**:1016–23. https://doi.org/10.1302/0301-620X.96B8.33766

47. Schünemann HJ BJ, Guyatt G, Oxman A, the GRADE Working Group. *GRADE Handbook*. 2013. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html (accessed 14 July 2017).

48. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Tugwell P. Developing core outcome sets for clinical trials: issues to consider. *Trials* 2012;**13**:132. https://doi.org/10.1186/1745-6215-13-132

49. Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CF, Askham J, Marteau T. Consensus development methods, and their use in clinical guideline development. *Health Technol Assess* 1998;**2**(3).

50. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials* 2007;**8**:38. https://doi.org/10.1186/1745-6215-8-38

51. Laisné F, Lecomte C, Corbière M. Biopsychosocial predictors of prognosis in musculoskeletal disorders: a systematic review of the literature (corrected and republished). *Disabil Rehabil* 2012;**34**:1912–41. https://doi.org/10.3109/09638288.2012.729362

52. van Rijn RM, Willemsen SP, Verhagen AP, Koes BW, Bierma-Zeinstra SM. Explanatory variables for adult patients' self-reported recovery after acute lateral ankle sprain. *Phys Ther* 2011;**91**:77–84. https://doi.org/10.2522/ptj.20090420

53. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med* 2011;**30**:377–99. https://doi.org/10.1002/sim.4067

54. White IR, Daniel R, Royston P. Avoiding bias due to perfect prediction in multiple imputation of incomplete categorical variables. *Comput Stat Data Anal* 2010;**54**:2267–75. https://doi.org/10.1016/j.csda.2010.04.005

55. Albert A, Anderson JA. On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* 1984;**71**:1–10. https://doi.org/10.1093/biomet/71.1.1

56. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;**3**:143–52. https://doi.org/10.1002/sim.4780030207

57. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361–87. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4

58. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;**49**:1373–9. https://doi.org/10.1016/S0895-4356(96)00236-3

59. Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, *et al.* Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLOS Med* 2014;**11**:e1001744. https://doi.org/10.1371/journal.pmed.1001744

60. Pavlou M, Ambler G, Seaman SR, Guttmann O, Elliott P, King M, Omar RZ. How to develop a more accurate risk prediction model when there are few events. *BMJ* 2015;**351**:h3868. https://doi.org/10.1136/bmj.h3868

61. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;**165**:710–18. https://doi.org/10.1093/aje/kwk052

62. Collins GS, Ogundimu EO, Cook JA, Manach YL, Altman DG. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. *Stat Med* 2016;**35**:4124–35. https://doi.org/10.1002/sim.6986

63. Royston P, Altman DG. Regression using fractional polynomials of continuous covariates: parsimonious parametric modelling. *J R Stat Soc Ser C Appl Stat* 1994;**43**:429–67. https://doi.org/10.2307/2986270

64. Royston P, Sauerbrei W. MFP. Multivariable Model-Building with Fractional Polynomials. In Royston P, Sauerbrei W, editors. *Multivariable Model-Building: A Pragmatic Approach to Regression Analysis Based on Fractional Polynomials for Modelling Continuous Variables*. Chichester: John Wiley & Sons, Ltd; 2008. pp. 115–50. https://doi.org/10.1002/9780470770771.ch6

65. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* 2007;**26**:5512–28. https://doi.org/10.1002/sim.3148

66. Marshall A, Altman DG, Holder RL, Royston P. Combining estimates of interest in prognostic modelling studies after multiple imputation: current practice and guidelines. *BMC Med Res Methodol* 2009;**9**:57. https://doi.org/10.1186/1471-2288-9-57

67. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 1987. https://doi.org/10.1002/9780470316696

68. Atkinson AC. A note on the generalized information criterion for choice of a model. *Biometrika* 1980;**67**:413–18. https://doi.org/10.1093/biomet/67.2.413

69. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making* 2006;**26**:565–74. https://doi.org/10.1177/0272989X06295361

70. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ* 2016;**352**:i6. https://doi.org/10.1136/bmj.i6

71. Thangaratinam S, Allotey J, Marlin N, Mol BW, Von Dadelszen P, Ganzevoort W, *et al.* Development and validation of Prediction models for Risks of complications in Early-onset Pre-eclampsia (PREP): a prospective cohort study. *Health Technol Assess* 2017;**21**(18). https://doi.org/10.3310/hta21180

72. Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol* 2016;**74**:167–76. https://doi.org/10.1016/j.jclinepi.2015.12.005

73. Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Stat Med* 2014;**33**:517–35. https://doi.org/10.1002/sim.5941

74. Wood AM, Royston P, White IR. The estimation and use of predictions for the assessment of model performance using large samples with multiply imputed data. *Biom J* 2015;**57**:614–32. https://doi.org/10.1002/bimj.201400004

75. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;**61**:76–86. https://doi.org/10.1016/j.jclinepi.2007.04.018

76. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;**35**:214–26. https://doi.org/10.1002/sim.6787

77. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, *et al.* Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010;**21**:128–38. https://doi.org/10.1097/EDE.0b013e3181c30fb2

78. Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, Woodward M. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart* 2012;**98**:691–8. https://doi.org/10.1136/heartjnl-2011-301247

79. Doherty C, Bleakley C, Hertel J, Caulfield B, Ryan J, Delahunt E. Recovery from a first-time lateral ankle sprain and the predictors of chronic ankle instability: a prospective cohort analysis. *Am J Sports Med* 2016;**44**:995–1003. https://doi.org/10.1177/0363546516628870

80. van Dijk CN, Lim LS, Bossuyt PM, Marti RK. Physical examination is sufficient for the diagnosis of sprained ankles. *J Bone Joint Surg Br* 1996;**78**:958–62. https://doi.org/10.1302/0301-620X78B6.1283

81. Guillo S, Bauer T, Lee JW, Takao M, Kong SW, Stone JW, *et al.* Consensus in chronic ankle instability: aetiology, assessment, surgical indications and place for arthroscopy. *Orthop Traumatol Surg Res* 2013;**99**(Suppl. 8):411–19. https://doi.org/10.1016/j.otsr.2013.10.009

82. Stiell IG, Greenberg GH, McKnight RD, Nair RC, McDowell I, Reardon M, *et al.* Decision rules for the use of radiography in acute ankle injuries. Refinement and prospective validation. *JAMA* 1993;**269**:1127–32. https://doi.org/10.1001/jama.269.9.1127

83. Bachmann LM, Kolb E, Koller MT, Steurer J, ter Riet G. Accuracy of Ottawa ankle rules to exclude fractures of the ankle and mid-foot: systematic review. *BMJ* 2003;**326**:417. https://doi.org/10.1136/bmj.326.7386.417

84. Doherty C, Bleakley C, Delahunt E, Holden S. Treatment and prevention of acute and recurrent ankle sprain: an overview of systematic reviews with meta-analysis. *Br J Sports Med* 2017;**51**:113–25. https://doi.org/10.1136/bjsports-2016-096178

85. Brison RJ, Day AG, Pelland L, Pickett W, Johnson AP, Aiken A, *et al.* Effect of early supervised physiotherapy on recovery from acute ankle sprain: randomised controlled trial. *BMJ* 2016;**355**:i5650. https://doi.org/10.1136/bmj.i5650

Appendix 1 Dynamic consent

SPRAINED study pilot of dynamic consent

Aim: To pilot dynamic consent in the SPRAINED study to explore how it might improve the consent procedure, and whether or not it influences trial adherence.

Objective: To determine whether or not dynamic consent can be introduced to a clinical study and integrated appropriately with study management software and existing recruitment processes.

Background: Dynamic consent is an approach to informed consent that is designed to allow participants to have greater control over how their samples and data are used, to interact with the study team more easily and to receive updates on how the research is progressing. Participants receive access to a personal profile that allows them to review their consent decisions, to change their mind and to receive relevant information about the study.

Researchers at the HeLEX centre have developed software to support a dynamic consent approach and have worked with members of the SPRAINED study research team to trial the software (tailored to the study) in the SPRAINED study, to see if it would influence trial retention rates. If participants were reminded of their involvement in the SPRAINED study, received notifications of upcoming questionnaires and were informed of the value of their continued involvement, even if they had fully recovered, it was hoped that this would help study retention.

It was important to ensure that dynamic consent did not adversely affect the SPRAINED study. On this basis, it was introduced in the later stages of recruitment once the centres had initiated recruitment processes and were familiar with the study. Ethics approval for the amendment to the study protocol was received, allowing dynamic consent to be implemented. Participants were consented if they visited the ED with a sprained ankle. The consent process in the case of the participants who were asked to trial dynamic consent was the same as for those following a traditional consent pathway, with an additional question included on the form asking whether or not they would be happy to use dynamic consent. They then signed a paper consent form, providing an e-mail address, and were sent a weblink to their secure dynamic consent page, where they could review their consent decisions or make any changes at any stage in the study. They also received notifications of any updates to the pages, including articles reminding them to complete the follow-up questions at 4 weeks, 4 months and 9 months.

Challenges: Dynamic consent presented a minor change to the recruitment process. As a result, implementing the change took longer than anticipated, as recruitment teams had to update their paperwork and remember to ask about involvement in the additional aspect of the study. Not all participants provided e-mail addresses, which limited the opportunity to set up dynamic consent accounts.

Results: Out of a total of 682 participants in the SPRAINED study, 22 were recruited to use dynamic consent. Of these 21 users, eight accessed their dynamic consent pages during the study (none of the participants changed their consent decisions during the study). It is not possible to determine from this whether or not dynamic consent improved response rates or study adherence; however, it was successful in demonstrating the possibility for dynamic consent software to integrate with clinical trial management software, and confirmed that the process for consent by using the dynamic consent software worked within a clinical setting.

Future work: Having confirmed the viability of the software, it is now important to apply it to a larger study, with a greater number of participants to further explore user experience, and to demonstrate how dynamic consent influences study experience and adherence.

Appendix 2 Systematic review search strategy

Allied and Complementary Medicine via OVID

Dates searched: 1985 to September 2016.

Date searched: 27 July 2016.

1. exp Ankle/
2. ankle.ti,ab.
3. Calcaneus/
4. calcane$.ti,ab.
5. Talus/
6. talus.ti,ab.
7. talocrural.ti,ab.
8. talofibular.ti,ab.
9. calcaneofibular.ti,ab.
10. Ankle Joint/
11. (ankle adj joint$).ti,ab.
12. Tarsal Joint/
13. (tarsal adj joint$).ti,ab.
14. Tarsal bones/
15. (tarsal adj bone$).ti,ab.
16. (lateral adj1 ligament$).ti,ab.
17. OR/1–16
18. Ankle Injury/
19. (ankle adj injur$).ti,ab.
20. Sprains and Strains/
21. (sprain$or strain$).ti,ab.
22. inversion.ti,ab.
23. OR/18–22
24. exp Prognosis/
25. prognos$.ti,ab.
26. predict$.tw.
27. exp Follow Up Studies/
28. (follow adj up adj stud$).ti,ab.
29. incidence.ti,ab.
30. course.ti,ab.
31. exp Longitudinal Studies/
32. longitudinal.ti,ab.
33. Prospective Studies/
34. prospect$.ti,ab.
35. Risk factors/
36. (risk adj factor$).ti,ab.
37. Cohort Studies/
38. (cohort adj stud$).ti,ab.
39. OR/24–38
40. 17 AND 23 AND 39

CENTRAL via EBSCO*host*

Dates searched: 1985 to September 2016.

Date searched: 26 July 2016.

#1 Ankle:MH 1364

#2 ankle:TI,AB,KY 4530

#3 (Ankle Joint):MH 505

#4 (ankle joint*):TI,AB,KY 814

#5 (Tarsal Bones):MH 16

#6 (tarsal bones):TI,AB,KY 19

#7 (tarsal joint*):TI,AB,KY 12

#8 (Tarsal Joints):MH 10

#9 Calcaneus:MH 115

#10 calcane*:TI,AB,KY 353

#11 Talus:MH 20

#12 talocrural:TI,AB,KY 25

#13 talofibular:TI,AB,KY 9

#14 calcaneofibular:TI,AB,KY 10

#15 (Lateral Ligament, Ankle):MH 0

#16 (lateral ligament*):TI,AB,KY 96

#17 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 4835

#18 (Ankle Injury):MH 0

#19 (ankle injur*):TI,AB,KY 561

#20 (Ankle Sprain):MH 0

#21 (ankle sprain):TI,AB,KY 245

#22 (Sprains and Strains):MH 267

#23 (sprain* or strain*):TI,AB,KY 7127

#24 inversion:TI,AB,KY 582

#25 #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 7923

#26 Prognosis:MH 10,961

#27 prognos*:TI,AB,KY 23,331

#28 Forecasting:MH 463

#29 predict*:TI,AB,KY 51,680

#30 (Follow Up):MH 48,086

#31 follow?up*:TI,AB,KY 2075

#32 Incidence:MH 7849

#33 incidence:TI,AB,KY 59,777

#34 (Cohort Studies):MH 6214

#35 (cohort stud*):TI,AB,KY 9473

#36 (Prospective Studies):MH 73,954

#37 (prospect* stud*):TI,AB,KY 97,763

#38 (Retrospective Studies):MH 6414

#39 (retrospect* stud*):TI,AB,KY 8809

#40 (Longitudinal Studies):MH 4966

#41 (longitudinal stud*):TI,AB,KY 6982

#42 (Risk Factors):MH 19,329

#43 (risk factor*):TI,AB,KY 35,375

#44 (Decision Support Techniques):MH 469

#45 #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 251,623

#46 #17 AND #25 AND #45 324

#47 fracture:TI,AB,KY 7565

#48 #17 AND #25 AND #45 NOT 47 302

#49 01/01/2015 TO 27/07/2016:CD 118,692

#50 #48 AND #49 33

Cumulative Index to Nursing and Allied Health Literature (CINAHL) via EBSCO*host*

Dates searched: 1982 to September 2016.

Date searched: 27 July 2016.

MH Ankle

TI ankle* OR AB ankle*

TI calcaneofibular OR AB calcaneofibular

TI talofibular OR AB talofibular

TI talocrural OR AB talocrural

TI (ankle N1 joint*) OR AB (ankle N1 joint*)

TI "tarsal joint*" OR AB "tarsal joint*"

TI "tarsal bone*" OR AB "tarsal bone*"

MH Calcaneus

MH Talus

MH Tarsal Bones+

MH Lateral Ligament, Ankle

TI (lateral N1 ligament) OR AB (lateral N1 ligament)

MH Ankle Sprain

MH Sprains and Strains

TI sprain* OR AB sprain*

TI strain* OR AB strain*

MH Ankle Injuries

TI (injur* N1 ankle) OR AB (injur* N1 ankle)

TI (inversion N1 sprain*) OR AB (inversion N1 sprain*)

MH Incidence

TI predict* OR AB predict*

TI "cohort stud*" OR AB "cohort stud*"

TI course OR AB course

MH Predictive research

MH Prognosis

TI prognos* OR AB prognos*

TI "follow up stud*" OR AB "follow up stud*"

TI "follow-up stud*" OR AB "follow-up stud*"

MH Prospective studies+

TI "longitudinal stud*" OR AB "longitudinal stud*"

MH Risk Factors

TI recovery OR AB recovery

TI (treatment N1 outcome*) OR AB (treatment N1 outcome*)

S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12 OR S13

S14 OR S15 OR S16 OR S17 OR S18 OR S19 OR S20

S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34

S35 AND S36 AND S37 retrieved 194 articles/204 articles on 26 July 2016

EMBASE via Ovid

Dates searched: 1974 to September 2016 week 30.

Date searched: 27 July 2016.

exp Ankle/

ankle.ti,ab.

Ankle Lateral Ligament/

(ankle adj lateral adj ligament).ti,ab.

Calcaneus/

calcane$.ti,ab.

Talus/

talus.ti,ab.

calcaneofibular.ti,ab.

talofibular.ti,ab.

talocrural.ti,ab.

(ankle adj joint$).ti,ab.

Tarsal Joint/

(tarsal adj joint$).ti,ab.

OR/1–14

Ankle Sprain/

Sprain/

sprain$.ti,ab.

strain$.ti,ab.

(inversion adj sprain$).ti,ab.

Ankle Injury/

OR/16–21

follow-up.mp.

prognos:.tw.

ep.fs.

OR/23–25

15 AND 22 AND 26

OpenGREY search strategy

Dates searched: from onset of database to September 2016.

Date searched: 27 July 2016.

Simple search in titles and abstracts for "ankle sprain or ankle"

Physiotherapy Evidence Database (PEDro) search strategy

Dates searched: onset of database to September 2016.

Date searched: 27 July 2016.

Simple search in titles and abstracts for "ankle sprains"

PsycINFO via Ovid

Dates searched: 1806 to July 2016 week 3.

Date searched: 27 July 2016.

1. exp Ankle/
2. ankle.ti,ab.
3. (ankle adj lateral adj ligament).ti,ab.
4. calcane$.ti,ab.
5. talus.ti,ab.
6. calcaneofibular.ti,ab.
7. talofibular.ti,ab.
8. talocrural.ti,ab.
9. (ankle adj joint$).ti,ab.
10. (tarsal adj joint$).ti,ab.
11. OR/1–10
12. sprain$.ti,ab.
13. strain$.ti,ab.
14. inversion.ti,ab.
15. OR/12–14
16. Prognosis/
17. prognos$.ti,ab.
18. predict$.ti,ab.
19. Followup Studies/
20. (follow?up adj stud$).ti,ab.
21. incidence.ti,ab.
22. course.ti,ab.
23. Longitudinal Studies/
24. (longitudinal adj stud$).ti,ab.
25. Prospective Studies/
26. (prospective adj stud$).ti,ab.
27. Risk Factors/
28. (risk adj factor$).ti,ab.
29. Cohort Analysis/
30. (cohort adj stud$).ti,ab.
31. Disease course/
32. OR/16–32

PubMed search strategy

Dates searched: onset of database to September 2016.

Date searched: 26 July 2016.

1. Ankle [mh] 2. ankle* [tiab] 3. Lateral Ligament, Ankle [mh] 4. calcane* [tiab] 5. Ankle Joint [mh] 6. ankle joint* [tiab] 7. tarsal joint* [tiab] 8. calcaneofibular [tiab] 9. talofibular [tiab] 10. talocrural [tiab] 11. talus [tiab] 12. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 13. Ankle Injuries [mh] 14. sprain* [tiab] 15. strain* [tiab] 16. Sprains and Strains [mh] 17. inversion [tiab] 18. #14 OR #15 OR #16 OR #17

19. Prognosis [MeSH:noexp] 20. diagnosed [tiab] 21. cohort* [tiab] 22. Cohort effect [mh] 23. Cohort studies [MeSH:noexp]

24. predictor* [tiab] 25. death [tiab] 26. "models, statistical" [mh]

27. #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26

28. #12 AND #18 AND #27

SportDiscus via EBSCO*host*

Dates searched: 1966–2016.

Date searched: 26 July 2016.

1. SU Ankle
2. TI ankle* OR AB ankle*
3. TI calcaneofibular OR AB calcaneofibular
4. TI talofibular OR AB talofibular
5. TI talocrural OR AB talocrural
6. TI "ankle joint*" OR AB "ankle joint*"
7. TI "tarsal joint*" OR AB "tarsal joint*"
8. TI "tarsal bones" OR AB "tarsal bones"
9. TI calcane* OR AB calcane*
10. TI talus OR AB talus
11. SU Ankle Lateral Ligament
12. TI "lateral ligament" OR AB "lateral ligament"
13. S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9 OR S10 OR S11 OR S12
14. SU Sprains
15. SU Strain
16. TI sprain* OR AB sprain*
17. TI strain* OR AB strain*
18. TI (injur* N1 ankle) OR AB (injur* N1 ankle)
19. TI (inversion N1 sprain*) OR AB (inversion N1 sprain*)
20. S12 OR S13 OR S14 OR S15 OR S16 OR S17 OR S18 OR S19
21. TI incidence OR AB incidence
22. TI predict* OR AB predict*
23. TI course OR AB course
24. TI cohort* OR AB cohort*
25. TI "cohort stud*" OR AB "cohort stud*"
26. SU Prognosis
27. TI prognos* OR AB prognos*
28. TI "follow up stud*" OR AB "follow up stud*"
29. TI "follow-up stud*" OR AB "follow-up stud*"
30. TI "longitudinal stud*" OR AB "longitudinal stud*"
31. TI "risk factor*" OR AB "risk factor*"
32. TI forecasting OR AB forecasting
33. TI "decision making" OR AB "decision making"
34. TI predict* and AB predict*
35. SU Cohort analysis
36. S21 OR S22 OR S23 OR S24 OR S25 OR S26 OR S27 OR S28 OR S29 OR S30 OR S31 OR S32 OR S33 OR S34 OR S35
37. S11 AND S20 AND S36

Appendix 3 Consensus meeting pre-meeting questionnaire

Please enter your Name here:__

Below is a questionnaire we would like you to complete and return prior to the Consensus Meeting on March the 27th. The results will inform our discussions during the meeting.
THEREFORE PLEASE COMPLETE AND RETURN BY WEDNESDAY 25TH MARCH

We have formatted the questionnaire so it is easiest to complete electronically. Once you have completed it, please save the file and include your surname in the file name and then email it back to us at sprained@ndorms.ox.ac.uk .

The questionnaire asks about different factors that may help predict recovery following an ankle sprain. Before you complete this you should look at the information provided in the summary pack that accompanies this questionnaire.

Your responses will be collated with those from other people attending the Consensus Meeting. During the meeting the overall group ratings will be summarised and you will have your own results provided to you in confidence for you to compare.
Please note there are no right or wrong answers.

You will be asked to respond to the questions using a nine point scale. In all cases please mark your response clearly in one box only. If you are completing this electronically, you just need to click on one box. An example is shown below:

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☒	☐	☐	☐	☐
	Not important				Important			Critical

In some cases you may feel you are unable to answer the question. In those cases please mark the "Don't know" box.

At the time of assessment in A&E, how important are the following factors in predicting recovery from an ankle sprain:

1. The time between injury and visiting A&E

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

---

2. The amount of ankle pain a person has

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

---

3. The amount of ankle pain a person has when putting weight on their injured ankle

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

---

4. The ability to put full weight on their ankle

1 2 3 4 5 6 7 8 9

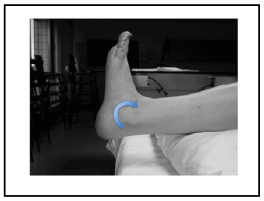
☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

5. The amount of ankle movement a person has pulling their toes up towards their head (dorsiflexion)



1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

6. The amount of ankle movement a person has pointing their toes away from their head (plantarflexion)

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

7. Abnormal imaging findings (for example ultrasound or MRI scans)

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

8. A person's age

1 2 3 4 5 6 7 8 9

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

Not important Important Critical

☐

Don't know

9. A person's Body Mass Index (combination of their weight and height)

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐
	Not important				Important			Critical

☐

Don't know

10. A person's working status (unemployed or working part-time or full time)

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐
	Not important				Important			Critical

☐

Don't know

11. A person's level of education

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐
	Not important				Important			Critical

☐

Don't know

12. *How* a person injured their ankle

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐
	Not important				Important			Critical

☐

Don't know

13. That a person has repeatedly sprained their ankle before

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

14. Whether a person's ankle is catching or locking

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

We would be interested to hear about other factors that you think are important in predicting recovery from an ankle sprain. Please type/write the most important factors below (maximum 2) and rate their importance.

15. Extra Factor A. _

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

16. Extra Factor B. _

1	2	3	4	5	6	7	8	9
☐	☐	☐	☐	☐	☐	☐	☐	☐

Not important Important Critical

☐

Don't know

Some research studies have shown that it is beneficial to collect information after the initial visit to A&E. Collecting delayed information can often improve the accuracy of the prediction of how people will recover following an ankle sprain.

---

17. If we were to collect further information like this, how many weeks after the initial visit do you think we should collect this information?

1 week	2 weeks	3 weeks	4 weeks	5 weeks	6 weeks
☒	☐	☐	☐	☐	☐

☐

Don't know

---

18. How should we collect this information?

Hospital visit	Postal Questionnaire	Online Questionnaire	Telephone Questionnaire
☐	☐	☐	☐

☐

Don't know

If you have any additional comments, please add them below:

MANY THANKS FOR TAKING TIME TO COMPLETE THIS

Appendix 4 Emergency department clinical data set form

BASELINE CLINICAL DATA SET
REC Reference: 15/LO/0538

SPRAINED

SITE

ATTACH PATIENT STICKER

Tel No. _____

_____

Best time: _____

Email add: _____

__ __ *(Research nurse: if patient is suitable for the study and has been registered, please tear/fold here to anonymise prior to sending a copy to the sprained study office)* __ __ __ __ .

SPRAINED BASELINE CLINICAL DATA SET: ANKLE INJURY (AGE 16 +YEARS)

HISTORY

1. Age: __ __ 2. Sex (tick): ☐ Male ☐ Female

3. Patient's reported Height: Feet __ __ Inches __ __ OR __ __ __cms

4. Patient's reported Weight: Stone __ __ lbs __ __ OR __ __ __ kgs

5. Currently employed (tick): ☐ None ☐ Part-time ☐ Full-time ☐ Student ☐ Retired

6. Days since injury __ __ Date of injury (DD/MM/YYYY) __ __/__ __/__ __ __ __

7. Injury setting (tick): ☐ At home ☐ At work/uni/school ☐ Playing sport ☐ Outside in public ☐ Other _____

8. Sprained this ankle in last 12 months (tick) ☐ Y ☐ N 9. Sprained this ankle at least twice before (tick) ☐ Y ☐ N

EXAMINATION

10. Ankle side (tick): ☐ L ☐ R 11. Patient able to weight bear (tick)? ☐ Y ☐ N

12. Pain at rest on 0-10 scale? (0= no pain, 10 = worst pain imaginable) __ __

13. Pain on weight bearing on 0-10 scale? (0= no pain, 10 = worst pain imaginable) __ __

14. Ankle movement limited (tick)? ☐ Y ☐ N

15. Since injury can patient dorsiflex fully? (circle one) Always / Often / Sometimes / Rarely / Never

16. Since injury can patient plantarflex fully? (circle one) Always / Often / Sometimes / Rarely / Never

17. Since injury has pt experienced catching/ locking when moving? (circle one) Always / Often / Sometimes / Rarely / Never

18. How long does patient expect recovery will take? (circle one)

< 2 wks / 2-8 wks / 2-6 mths / 6-12 mths / > 1 year / Not sure will recover / don't know

INVESTIGATION

Xray of ankle/foot: (tick) ☐ Y ☐ N

SPRAIN SEVERITY (tick)

Mild (Gd I) / Moderate (Gd II) / Severe (Gd III)

☐ ☐ ☐

SUITABLE FOR SPRAINED STUDY?

PATIENT CONTACT DETAILS (mobile) CONFIRMED AS CORRECT ☐

Is patient suitable for the SPRAINED Study(tick)? ☐ Y ☐ N If YES, trial information and invite given? ☐ Y ☐ N

If patient does not wish to be contacted about SPRAINED please tick here ☐

Record reason here if patient declines SPRAINED study_____

Signature of clinician:_____ Date form completed: ___/___/_____

PLEASE NOW HAND THIS FORM TO YOUR SPRAINED RESEARCH CLINICIAN / PLACE IN SPRAINED BOX FILE

When registered on to the SPRAINED Study: Enter study number and initials below, anonymise this sheet and post to the SPRAINED Study Office in Freepost envelope.

Study code:		Site ID code:		Participants Study Number:					Initials:		
S	P										

SPRAINED_BASELINE_V1.0_CORE_15JUN2015.docx Page 1 of 1

EME
HS&DR
HTA
PGfAR
PHR

Part of the NIHR Journals Library
www.journalslibrary.nihr.ac.uk

Published by the NIHR Journals Library